
PARA COMPRENDER A RAWLS
UNA RECONSTRUCCION Y UNA CRITICA DE LA
TEORIA DE LA JUSTICIA

Teoría de la justicia (FCE, 1979), el célebre ensayo de John Rawls, apareció en inglés en 1971 y de inmediato se convirtió en materia de un debate en el que no faltaron los malentendidos y las tergiversaciones. Hacía falta, por ello, una obra que esclareciera los postulados de Rawls y los expusiera y examinara con claridad. A esa finalidad está consagrado el libro de Robert Paul Wolff, y consigue sus propósitos con excelencia. Se trata, dicho de manera sumaria, de un texto de orden sociofilosófico: es al mismo tiempo una reconstrucción de Teoría de la justicia y un abordaje crítico. Wolff, profesor de filosofía de la Universidad de Massachusetts en Amherst, ha echado mano de múltiples recursos intelectuales para llevar a cabo su investigación, fruto vivo de sus cursos académicos. Al penetrar la sustancia del discutido libro de Rawls, muestra sus debilidades y revela sus riquezas; constituye un esfuerzo ejemplar por entender los problemas de la justicia en el mundo de nuestros días.

FONDO DE CULTURA ECONOMICA

ROBERT PAUL WOLFF PARA COMPRENDER A RAWLS



ROBERT PAUL WOLFF

PARA
COMPRENDER
A RAWLS

UNA RECONSTRUCCION
Y UNA CRITICA DE LA
TEORIA DE LA JUSTICIA



Traducción de
MARCIAL SUÁREZ

ROBERT PAUL WOLFF

PARA COMPRENDER A RAWLS

*Una reconstrucción y una crítica de
Teoría de la Justicia*



FONDO DE CULTURA ECONOMICA
MEXICO

Primera edición, 1981

PARA COMPRENDER A RAWLS

Una introducción y una crítica de
Tomas de la Parra

Título original:

Understanding Rawls

© 1977 Princeton University Press, Princeton, New Jersey

D. R. © 1981 FONDO DE CULTURA ECONÓMICA

Av. de la Universidad, 975; México 12, D. F.

ISBN 968-16-0792-9

Impreso en México

A la memoria de
Charlotte Ornstein Wolff
1900-1973

Este libro es el resultado de una colaboración por décadas entre dos de los filósofos más importantes de la filosofía política del siglo XX. El profesor John Rawls, de la Universidad de Princeton, y el profesor Charles L. Stevenson, de la Universidad de California en Berkeley, han trabajado juntos durante más de treinta años en la elaboración de este libro. El libro es el resultado de una colaboración que ha sido fructífera y que ha producido una obra que es una contribución importante a la filosofía política.

El libro comienza con una introducción de John Rawls, en la que se discute la importancia de la filosofía política y la necesidad de una teoría de la justicia. Después, el libro se divide en tres partes. La primera parte, "La teoría de la justicia", discute la teoría de la justicia de Rawls y su aplicación a la filosofía política. La segunda parte, "La teoría de la justicia en la práctica", discute la aplicación de la teoría de la justicia a la filosofía política. La tercera parte, "La teoría de la justicia en la historia", discute la aplicación de la teoría de la justicia a la historia.

El libro es una obra importante que contribuye a la filosofía política y a la comprensión de la teoría de la justicia de Rawls. Es una obra que es esencial para cualquier estudiante de filosofía política y para cualquier filósofo interesado en la filosofía política.

Este libro es el resultado de una colaboración por décadas entre dos de los filósofos más importantes de la filosofía política del siglo XX. El profesor John Rawls, de la Universidad de Princeton, y el profesor Charles L. Stevenson, de la Universidad de California en Berkeley, han trabajado juntos durante más de treinta años en la elaboración de este libro. El libro es el resultado de una colaboración que ha sido fructífera y que ha producido una obra que es una contribución importante a la filosofía política.

Nottingham
Mayo de 1978

PREFACIO

ESTE LIBRO es el resultado de mis esfuerzos por clarificar *Teoría de la Justicia* en mi propio interés y en el de mis alumnos de un curso sobre filosofía política en la Universidad de Massachusetts, en Amherst. Cuando me disponía a preparar mis lecciones acerca de la filosofía de John Rawls, en el otoño de 1975, sentí la necesidad de escribir mis observaciones a fin de mantener mis ideas en orden. Esas observaciones constituyeron el manuscrito que sirvió de origen al presente libro.

Un buen número de alumnos formuló puntos de vista, razonamientos, objeciones y sugerencias que me fueron muy útiles. Judith Decew, en su trabajo de fin de curso, llamó mi atención sobre la exposición de John Harsanyi acerca de las asignaciones de *maximin* y de probabilidad, que yo he resumido en una nota en el capítulo XV. Peter Markie me ayudó a comprender algunas de las diferencias lógicas entre las formas primera y última del juego del regateo de Rawls. Y Jim Lane y Mark Richard, en el curso de un seminario sobre la "teoría del juego", desarrollado durante la primavera de 1976, mejoraron enormemente mi interpretación de la estructura lógica del "juego del regateo" bajo el "velo de la ignorancia".

De acuerdo con lo que en mí es ya costumbre —y muy grata, por cierto—, consulté a mi buen amigo y colega, el profesor Robert J. Ackermann, sobre algunas cuestiones técnicas que amenazaban con desbordarme.

Sólo quienes se hallen familiarizados con la crítica literaria de mi mujer, la profesora Cynthia Griffin Wolff, reconocerán toda la influencia que esa crítica ha ejercido sobre mi trabajo. A lo largo de muchos años, mi mujer ha profundizado y dado mayor complejidad a mi conocimiento del espíritu del hombre y de la palabra escrita. A la vez que preparaba para la publicación su libro más reciente, ordenó su tiempo para poder leer cuidadosamente este manuscrito. El resultado fue la eliminación de un cierto número de notables errores e impropiedades. Este ensayo, como toda mi filosofía, está escrito para ella.

Norhampton
Mayo de 1976

El presente libro es el resultado de una investigación que comenzó en el año 1974, cuando el autor, como profesor de la asignatura de Historia del Arte en la Universidad de Valencia, se planteó el problema de la evolución de la pintura española durante el siglo XVIII. En ese momento, la investigación se centró en el estudio de la obra de los pintores de la escuela de Valencia, pero a medida que avanzaba, se fue ampliando el campo de estudio para abarcar a todos los pintores de la época. El libro está dividido en tres partes: la primera trata de la evolución de la pintura española durante el siglo XVIII; la segunda trata de la obra de los pintores de la escuela de Valencia; y la tercera trata de la obra de los pintores de la escuela de Madrid. El libro está escrito en un lenguaje claro y sencillo, y está acompañado de una gran cantidad de reproducciones de obras de arte.

El autor quiere agradecer a todos los que han colaborado en esta investigación, especialmente a los profesores de la Universidad de Valencia, a los amigos y a la familia. También quiere agradecer a los editores de esta obra, a los que han permitido que esta investigación sea conocida por un mayor número de personas.

El presente libro es el resultado de una investigación que comenzó en el año 1974, cuando el autor, como profesor de la asignatura de Historia del Arte en la Universidad de Valencia, se planteó el problema de la evolución de la pintura española durante el siglo XVIII. En ese momento, la investigación se centró en el estudio de la obra de los pintores de la escuela de Valencia, pero a medida que avanzaba, se fue ampliando el campo de estudio para abarcar a todos los pintores de la época. El libro está dividido en tres partes: la primera trata de la evolución de la pintura española durante el siglo XVIII; la segunda trata de la obra de los pintores de la escuela de Valencia; y la tercera trata de la obra de los pintores de la escuela de Madrid. El libro está escrito en un lenguaje claro y sencillo, y está acompañado de una gran cantidad de reproducciones de obras de arte.

El presente libro es el resultado de una investigación que comenzó en el año 1974, cuando el autor, como profesor de la asignatura de Historia del Arte en la Universidad de Valencia, se planteó el problema de la evolución de la pintura española durante el siglo XVIII. En ese momento, la investigación se centró en el estudio de la obra de los pintores de la escuela de Valencia, pero a medida que avanzaba, se fue ampliando el campo de estudio para abarcar a todos los pintores de la época. El libro está dividido en tres partes: la primera trata de la evolución de la pintura española durante el siglo XVIII; la segunda trata de la obra de los pintores de la escuela de Valencia; y la tercera trata de la obra de los pintores de la escuela de Madrid. El libro está escrito en un lenguaje claro y sencillo, y está acompañado de una gran cantidad de reproducciones de obras de arte.

El autor quiere agradecer a todos los que han colaborado en esta investigación, especialmente a los profesores de la Universidad de Valencia, a los amigos y a la familia. También quiere agradecer a los editores de esta obra, a los que han permitido que esta investigación sea conocida por un mayor número de personas.

El presente libro es el resultado de una investigación que comenzó en el año 1974, cuando el autor, como profesor de la asignatura de Historia del Arte en la Universidad de Valencia, se planteó el problema de la evolución de la pintura española durante el siglo XVIII. En ese momento, la investigación se centró en el estudio de la obra de los pintores de la escuela de Valencia, pero a medida que avanzaba, se fue ampliando el campo de estudio para abarcar a todos los pintores de la época. El libro está dividido en tres partes: la primera trata de la evolución de la pintura española durante el siglo XVIII; la segunda trata de la obra de los pintores de la escuela de Valencia; y la tercera trata de la obra de los pintores de la escuela de Madrid. El libro está escrito en un lenguaje claro y sencillo, y está acompañado de una gran cantidad de reproducciones de obras de arte.

OBSERVACIONES PRELIMINARES

*TEORÍA DE LA JUSTICIA**, de John Rawls, es un libro importante, pero es un libro desconcertante también. Es extremadamente largo, y algunas partes de él se desarrollan lentamente. Rawls pasa, repetidamente, del más elaborado despliegue de los modelos formales de la economía y de la matemática a discusiones de temas anticuados, así como a materias y referencias tomadas de las escuelas utilitarista, utópica, intuicionista y psicológica empírica del pensamiento inglés que floreció a finales del siglo XIX y comienzos del XX. El libro muestra todos los indicios de haber sido minuciosamente anotado, unificado y sintetizado, como si cada elemento de la exposición hubiera sido valorado en relación con los demás; pero hay numerosas y graves incoherencias y confusiones que producen la impresión de que Rawls no ha podido decidirse acerca de algunas cuestiones muy fundamentales. El carácter lógico de los objetivos del libro nunca resulta enteramente claro, a pesar del evidente interés de Rawls por cuestiones de este tipo. En muchos sitios parece admitir, sencillamente, que ha ajustado sus premisas de tal modo que desemboquen en las conclusiones que él desea; pero, por todas partes, desde las primeras páginas hasta las últimas, pretende haber demostrado, o, por lo menos, haber esbozado la demostración de algo parecido a un "teorema".

La importancia de la obra de Rawls, y también su falta de claridad, están confirmadas por la abundancia de comentarios que han aparecido en los cinco años transcurridos desde su publicación. Filósofos, economistas y especialistas en ciencias políticas han tratado de acercarse al libro, con resultados, a veces, similares a los de unos ciegos que intentasen alcanzar a un elefante¹. Yo creo que es posible captar una

* Fondo de Cultura Económica, 1979.

¹ Han aparecido ya varias bibliografías de libros y artículos sobre Rawls, aunque todas ellas se quedan anticuadas poco después de su publicación. Los lectores interesados en examinar esta literatura encontrarán de cierta utilidad una breve bibliografía anotada al final de este ensayo. Cuando mis observaciones recogen las de otros autores, he procurado indicarlo en el texto.

gran parte del sentido del libro de Rawls, aunque, en última instancia, no me parezca que sus objetivos se hayan cumplido. Pero, para captar ese sentido, tenemos que adoptar una actitud exegética un tanto ajena a la ortodoxia.

En resumen, lo que yo propongo es una lectura de *Teoría de la Justicia*, no como una pieza única de contenido filosófico que ha de someterse a examen para ser aceptada o rechazada en su conjunto, sino como un complejo y múltiple informe de veinte años, por lo menos, de crecimiento y desarrollo filosófico. Señalaré que Rawls comenzó con un problema simple, coherente, comprensible, y con una brillante idea para su solución. Su propósito original debe de haber sido el de escribir un libro muy semejante al de Kenneth Arrow, *Social Choice and Individual Values*, que presentase la solución de su problema como un teorema formal de enorme poder y significación. La idea no llegó a surtir efecto, a pesar de su gran originalidad. Las laberínticas complejidades de *Teoría de la Justicia* son las consecuencias de tres fases, por lo menos, en el desarrollo del pensamiento de Rawls, en cada una de las cuales el autor complicaba su teoría para resolver las objeciones que otros le habían planteado a la vista de las versiones anteriores, o las que él mismo percibía. En un momento clave de ese desarrollo —aproximadamente, cuando introdujo la noción del “velo de la ignorancia”—, Rawls descubrió una forma de relacionar su elaboración con una tradición totalmente distinta de la teoría moral y política, y este hecho, evidentemente, le impulsó a creer que la versión revisada y complicada era superior a la original. A medida que se agregaban advertencias, matices y complicaciones a la teoría, iba resultando menos aceptable intuitivamente, y mucho menos semejante a un “teorema” del tipo que Rawls se proponía demostrar. Pero el lenguaje de su libro revela que nunca abandonó la ilusión, utilizando para ese fin términos como “prueba”, “teorema” y “teoría”.

Mi reconstrucción histórica del desarrollo de la teoría de Rawls no es enteramente especulativa, desde luego. Lo que yo llamaré la primera forma del modelo fue expuesto en su ensayo titulado “La justicia como imparcialidad”, que apareció en la *Philosophical Review*, en 1958, y que ha tenido frecuentes reimpresiones². Una versión muy distinta de la teoría, correspondiente, en líneas generales, a lo que yo llamo la segunda forma del modelo, fue la ofrecida por Rawls, en

² Me referiré a este ensayo, tal como aparece en *Philosophy, Politics, and Society*, dirigido por Peter Laslett y W. G. Runciman (Barnes & Noble, Inc., 1962), segunda serie.

1967, en su artículo “Justicia distributiva”, que apareció como ensayo original en la tercera serie de *Philosophy, Politics, and Society*, de Laslett y Runciman. Estos dos ensayos, junto a la versión final de *Teoría de la Justicia*, constituyen una importante muestra del desarrollo del pensamiento de Rawls.

De todos modos, el principal argumento en favor de la reconstrucción que voy a desarrollar en este ensayo es su utilidad a la hora de ayudarnos a comprender un buen número de elementos de la teoría final, cuya presencia y cuya función precisa serían, de otro modo, ampliamente desorientadoras. Gracias a esta reconstrucción, podremos comprender por qué se supone, más bien arbitrariamente, que los participantes en la posición original están libres de envidia; por qué hay que formular una teoría difícil y polémica de los bienes primarios; por qué es necesario un velo de la ignorancia, y por qué debe suponerse que no se borra el conocimiento de lo que Rawls llama, de un modo más bien vago, los hechos básicos de la sociedad y de la psicología moral; por qué se asegura que el primer principio de justicia alcanza prioridad sobre el segundo; y por qué los participantes en la posición original esquivan, un tanto extrañamente, el principio de razón insuficiente, en favor del altamente polémico principio maximin de adopción de decisiones en condiciones de incertidumbre.

En el desarrollo de mi reconstrucción, antes de alcanzar el punto en que formulé mis propias críticas de la teoría, centraré la atención en algunos elementos de los diversos “modelos” que Rawls ha alterado posteriormente o abandonado por completo; e ignoraré o trataré más bien sumariamente otros elementos de la teoría final a los que Rawls concede, explícitamente, gran importancia. Esta forma de crítica e interpretación suscita algunas difícilísimas cuestiones de método filosófico, y acaso lo mejor sea poner mis cartas sobre la mesa en las primeras páginas de este ensayo, a fin de que los lectores puedan juzgar mejor lo que sigue. Al propio tiempo, espero que lograré explicar por qué creo que este ensayo puede ser útil y de interés, tanto para el que empieza a estudiar a Rawls y busca una guía que le permita dominar *Teoría de la Justicia*, como para el especialista o el crítico que quieran formarse algún criterio acerca de su duradera importancia y validez.

Hay entre los filósofos americanos contemporáneos dos concepciones contrarias del modo de hacer filosofía, de cuáles son los signos de la buena filosofía, y de la forma en que debe juzgarse, positiva o negativamente, el valor de las tesis y de los argumentos filosóficos. Aunque ambas concepciones son defendidas, de diversos modos y en distintos

grados, por filósofos que se describirían como "analíticos" en su metodología y en su orientación, encuentran expresión en muy diferentes tipos de obras filosóficas. Un punto de vista es el de que el valor de una posición filosófica consiste en la precisión, en el detalle y en la perfección con que está elaborada. Los filósofos de esta opinión dedican una gran cantidad de energía y de imaginación a defender sus planteamientos contra las objeciones que se les hacen, sobre todo cuando esas objeciones adoptan la forma de contra-ejemplos frente a las tesis generales que se han formulado. El carácter representativo del *cogito* de Descartes, o la noción kantiana del argumento transcendente, se someten a una disección anatómica, con sucesivas formulaciones y reformulaciones de la misma idea central, que se proponen, se critican, se revisan y se vuelven a criticar. El modelo de actividad filosófica de este tipo es la lógica formal, en la que un teorema es invalidado por un solo contra-ejemplo, aunque sea muy grotesco y periférico. Considero justo señalar que es probable que los defensores de esta posición metodológica prefieran la versión final a gran escala, increíblemente detallada, de la teoría de Rawls, tal como se expone en *Teoría de la Justicia*, a los esbozos de la misma, relativamente breves, que aparecieron en sus artículos anteriores.

La concepción opuesta de la filosofía, que yo comparto, es la de que el valor real de una posición filosófica radica casi enteramente en la profundidad, en la penetración y en la fuerza de su idea central. En la *Crítica de la razón pura*, por ejemplo, Kant desarrolla su estrategia de defensa de las pretensiones fundamentales de la ciencia y de la matemática contra el ataque escéptico, mostrándolas como basadas en la posibilidad de la conciencia en general. Nada auténticamente nuevo e importante ha sido agregado por los incontables filósofos que han clasificado, catalogado, criticado y multiplicado "argumentos trascendentes" en periódicos y revistas.

Una nueva idea filosófica puede ser, por supuesto perfectamente técnica. Al igual que la idea esencial de una demostración matemática, esa nueva idea filosófica puede requerir una considerable formación previa para ser comprendida. Pero, al igual que una idea matemática realmente original, puede también ser captada, por lo general, mediante un solo acto de pensamiento. Las tesis centrales de la *Crítica de la razón pura* son fundamentalmente sencillas —no fáciles de comprender, pero no elaboradamente complicadas en su detalle. Yo creo que la filosofía avanza por saltos de *quantum* a medida que se van alcanzando puntos de vista auténticamente nuevos; no avanza paso a paso, impul-

sada por la tenaz labor de innumerables pensadores profesionales.

¿Cuál es la relación de estas observaciones con este ensayo? Si usted se adscribe a la primera concepción de la filosofía, un libro sobre Rawls tiene que ser, o bien elemental, introductorio y expositivo, o bien, en otro caso, avanzado, detallado y crítico. Una "contribución al saber" consistirá en la crítica de determinados puntos de Rawls, cuya comprensión presupone que ya se tiene un sólido conocimiento del texto original. Pero si está usted de acuerdo conmigo en la adhesión a la segunda concepción de la filosofía, creará que es, por lo menos, posible escribir un ensayo sobre Rawls que sea, a la vez, una ayuda para su comprensión por parte del alumno inteligente que empieza, y también una auténtica contribución filosófica al saber en este campo. Para conseguir esto, tendremos que captar el núcleo central o idea de *Teoría de la Justicia*, ponerla al descubierto de una manera clara, y exponer luego sus virtudes y sus defectos como idea fundamental, independientemente de las particularidades en que se inserta. Si tenemos éxito, podremos comprender esas particularidades como variaciones sobre la idea central, como defensas de ésta contra posibles ataques, y así sucesivamente.

El orden de mi exposición será como sigue. En los dos capítulos restantes de la primera parte, esbozaré el problema original o el conjunto de problemas con que Rawls se enfrentaba cuando comenzó el desarrollo de su teoría. Después, expondré lo que yo considero que es la idea central o clave de la obra de Rawls, e indicaré por qué, en mi opinión, se trata de una idea filosófica original, importante y poderosa.

En la segunda parte, desarrollaré la evolución de la teoría de Rawls, desde su primera forma, relativamente simple, en el artículo de 1958, "La justicia como imparcialidad", hasta la complejidad barroca final de *Teoría de la Justicia*. Procederé dialécticamente, estableciendo la primera forma del modelo de Rawls, someténdola a crítica, pasando a la segunda forma revisada del modelo, someténdola, a su vez, a crítica, y estudiando luego los elementos de la forma a escala natural del modelo, tal como aparece en el libro. El propósito de este modo de exposición es el de demostrar que muchos de los elementos de la teoría final fueron introducidos, no por su valor filosófico intrínseco, sino, más bien, como recursos para salir al paso de reales o posibles objeciones. Una complicación particularmente importante del modelo final, la llamada "prioridad de la libertad", se incorporará entonces a la descripción, mediante su descomposición en distintos factores, y la se-

gunda parte se cerrará con una breve discusión de ciertas elaboraciones y complicaciones secundarias que Rawls añade al modelo en su forma oficial.

La tercera parte de este ensayo está dedicada a una amplia discusión de la relación entre Rawls y Kant, con especial atención al punto de vista del propio Rawls acerca de esa relación. Como veremos, Rawls está equivocado en lo que se refiere a las conexiones entre su filosofía política y la teoría moral de Kant, pero es perfectamente justo al insistir sobre el significado de la comparación.

Las tres primeras partes de este ensayo son esencialmente reconstructivas y expositivas, a pesar de la presencia de un considerable aparato de discusión crítica. La cuarta parte, construida sobre el análisis de las secciones precedentes, presenta mi crítica propiamente dicha de la teoría de Rawls. Algunas de mis observaciones repiten las objeciones de otros profesores, y, mediante referencias a pie de página, trataré de dar al lector alguna guía acerca de la literatura crítica existente; creo que algunas de mis críticas son originales, aunque la respuesta a Rawls ha sido tan rápida y extendida, que no puedo estar seguro de que alguien no se me haya adelantado. El propósito de la discusión de la cuarta parte será el de demostrar que el modelo de Rawls es, finalmente, insatisfactorio, a pesar de los ajustes más bien imaginativos con que él trata de apuntalarlo. Como Rawls inicia su libro con algunas observaciones metodológicas relativas a lo que él llama "equilibrio reflexivo", yo concluyo la cuarta parte con una discusión del *status* lógico del tema de *Teoría de la Justicia*.

Si yo cerrase mi discusión en este punto, seguiría en pie la cuestión más importante de todas: ¿qué hemos de hacer con la teoría de la justicia de Rawls? En el caso de que no me equivoque al considerar que el valor de la obra de un filósofo radica en el poder y en la fecundidad de su concepción central, más que en el detalle de su exposición, difícilmente podré soslayar la responsabilidad de formular algún juicio acerca de la idea que yo percibo en el corazón de la filosofía de Rawls. Creo que la confianza de Rawls en ciertos modelos formales de análisis sacados de la teoría de la elección racional es fundamentalmente errónea, y que su empleo de los conceptos y modelos de la teoría de la utilidad, de la economía del bienestar y de la teoría del juego, que está en el corazón mismo de su empresa, es un modo equivocado de abordar los problemas normativos y explicativos de la teoría social. En la última parte de este ensayo, trabajaré sobre este propósito, y haré todo lo posible por que resulte aceptable.

II

EL PROBLEMA

EL PROBLEMA con que Rawls comienza es el callejón sin salida de la teoría ética angloamericana, hacia principios de los años 1950. Si dejamos a un lado el emotivismo en todas sus diversas formas, las más importantes escuelas de conocimiento de la teoría ética eran el utilitarismo y el intuicionismo. Cada una de estas tradiciones tiene virtudes, desde el punto de vista de Rawls, pero cada una de ellas tiene también fatales defectos. Rawls revive una versión de la teoría del contrato social como un modo de descubrir una *vía media* entre el utilitarismo y el intuicionismo.

Las principales virtudes del utilitarismo son, en primer lugar, su terminante afirmación del valor fundamental de la felicidad humana, y, en segundo lugar, su carácter constructivo — es decir, su enunciación de una norma o procedimiento mediante el cual las cuestiones éticas encuentran una respuesta y las disputas éticas una solución. Un valor secundario del utilitarismo, tanto para sus creadores como para Rawls, es su aptitud como principio para el arreglo de cuestiones de política social. Los dos defectos más evidentes del utilitarismo son su imposibilidad de explicar el grado de racionalidad con que han de ser guiados los interesados en elevar al máximo su propio placer, a fin de que sustituyan, como objeto de sus acciones, su felicidad personal por la felicidad general, y las implicaciones manifiestamente contra-intuitivas, y, a veces, auténticamente aborrecibles, de su principio fundamental. Los ejemplos de esta segunda incapacidad son bien conocidos, naturalmente, pero vale la pena recordar que los más notables contra-ejemplos frente al principio de utilidad surgen, precisamente, en relación con cuestiones de justicia procesal o sustantiva.

Como teoría moral, el intuicionismo es metodológicamente inferior al utilitarismo. Afirma, simplemente, de un modo absoluto y sin demostración, que cada uno de nosotros tiene un poder de "intuición moral", llamado "racional" por los intuicionistas, pero que no muestra ninguna estructura de razonamiento práctico mediante la cual podamos aprender directamente la obligatoriedad de los actos particu-

res⁵. Ni siquiera la agregación por parte de W. D. Ross de la noción de la obligación *prima facie* es capaz de transformar el intuicionismo en una teoría de razonamiento moral, porque cuando nos vemos precisados a valorar la importancia relativa de los deberes *prima facie* que entran en oposición recíproca, tenemos que confiar, una vez más, solamente en la intuición. Pero, si bien el intuicionismo es débil como descripción de un razonamiento práctico, es fuerte en dos sentidos que son claramente importantes para Rawls. Primero: define lo justo independientemente de lo bueno, y hace así de la rectitud una noción fundamental, irreductiblemente moral. Segundo: toma de Kant la doctrina de la inviolabilidad y de la dignidad de la personalidad moral, y rechaza de este modo, decisivamente, la tendencia utilitarista a considerar a los seres humanos sólo como recipientes de placer, que han de llenarse o de vaciarse como otras tantas vasijas de agua.

En esta disputa, las simpatías y las antipatías de Rawls se hallan divididas por igual. Moralmente, él se encuentra más cómodo, sin duda, con los intuicionistas que con los utilitaristas; pero, metodológicamente, su corazón está con los utilitaristas y con los economistas neoclásicos que adoptaron el utilitarismo como fundamento moral de sus elegantes construcciones teóricas. El utilitarismo, incluso en sus versiones más elaboradas y complicadas, apoya el sacrificio de algunas personas a la felicidad de otras. Como Rawls sostiene en un buen número de secciones de su libro, el utilitarismo trata a los seres humanos individuales como otras tantas dimensiones a lo largo de las cuales puede distribuirse la felicidad, más que como agentes morales autónomos cada uno de los cuales sigue, independientemente, una vía de acción con razón y dignidad, libremente elegida. Pero el intuicionismo es una simple expresión de convicción, una confesión del fracaso de la razón práctica. Incluso al más honesto y bienintencionado agente moral que se pregunta lo que ha de hacer o que percibe un conflicto entre dos obligaciones reconocidas, el intuicionista sólo puede decirle: "consulta a tus intuiciones". Si mis intuiciones difieren de las del intuicionista, él no puede presentar, literalmente, ninguna argumentación para convencerme, por muy dispuesto que yo esté a escucharle.

Con este atolladero como punto de partida, Rawls define su problema. Podemos imaginarle pensando: "Empezaré con una concepción de la acción racional tan escrupulosa y moralmente neutral como

⁵ Ver, por ejemplo, W. D. Ross, *The Right and The Good*, o H. A. Prichard, *Moral Obligation*.

pueda alcanzarse aceptablemente, concepción que sería admisible para los utilitaristas y también para los intuicionistas. Supondré que los hombres buscan la felicidad, pero no concluiré que la felicidad es El Bien, ni caeré, desde luego, en la suposición de que la felicidad de todos es El Bien para todos. Es evidente que la simple noción de acción racional es insuficiente para una moralidad racionalmente defendida, pero, ¿hay algo que pueda agregarse a estos elementos, además del puro planteamiento de sustantivas pretensiones morales, que proporcione un argumento racional para un principio moral que sea constructivo, que se ajuste a nuestras más sólidas convicciones morales —o que, por lo menos, no se desvíe excesivamente de ellas—, que dé una calificación lógica independiente a la noción de lo justo y que adopte como fundamental la inviolabilidad y la dignidad de la personalidad moral?"

Llegado a este punto, Rawls se vuelve a una tercera tradición, no precisamente de filosofía moral o social, sino más bien de teoría política: la tradición del contrato social. Históricamente, la idea del contrato social ha sido invocada con el fin de explicar o justificar la autoridad del estado. Perteneció a la tradición, más larga todavía, de los debates acerca de la naturaleza y de la sede de la soberanía, y se interesa, en consecuencia, por los problemas formales de la naturaleza y de la legitimidad de la autoridad, más que por los problemas sustantivos de la política social y de la justicia distributiva. El tema común de las diversas teorías del contrato es la aseveración de que la soberanía reside originalmente en los individuos, de modo que la reunión o la agregación de la misma en una unidad y el traslado de ella a un portador que ha de ejercerla requiere un acuerdo colectivo (y generalmente unánime) de los originales poseedores de la soberanía.

Desde el punto de vista de la teoría moral, hay, en realidad, dos tradiciones en torno al contrato. La primera, representada por Locke, supone una teoría moral como una de las premisas del razonamiento relativo al contrato. En el *Segundo tratado* de Locke, se formula, simplemente, una teoría moral a gran escala de los derechos naturales y de la ley natural, aproximadamente a lo largo de líneas intuicionistas. Según Locke, la sustancia y la validez del contrato descansan en esa teoría, que le dice, no sólo por qué el estado tiene derecho a dominar y cuáles son sus legítimos objetivos, sino también cuáles son los límites de su autoridad y en qué condiciones los individuos tienen derecho a derribarlo.

La segunda tradición relativa al contrato procede de Rousseau. Co-

mienza también con un estado de naturaleza y con una teoría de la naturaleza humana, pero afirma (aunque paradójicamente) que el ingreso en la sociedad civil por medio de un contrato social opera una transformación moral en los contratantes originales. El contrato cambia, no sólo la sustancia, sino también la forma del razonamiento moral de los hombres. Antes del acuerdo, cada individuo tiene lo que Rousseau llama "una voluntad privada". Pero, al establecer el convenio, cada individuo adquiere una voluntad general, o, mejor, adquiere la posibilidad por primera vez, de tener o de ejercer una voluntad general. La transformación es moral, no psicológica. La persona más benévola imaginable no podría, en un estado de naturaleza, tener una voluntad general, de igual modo que un grupo de colaboradores no podría, sin un sistema legal, constituirse en sociedad limitada.

Rawls propone superar el punto en que las teorías morales del utilitarismo y del intuicionismo se han atascado, invocando una versión de la teoría del contrato social en su forma rousoniana. Al hacerlo así, quiere unir la filosofía moral y la social, de un modo que recuerda más a Platón que a Locke, Bentham o Mill. Y espera que así podrá formular un principio fundamental de teoría moral y social que sea constructivo, racional, que atienda tanto al bien de la felicidad humana como a la dignidad de la personalidad moral, y mediante el cual pueda darse una argumentación que tenga alguna posibilidad de persuadir a los que no estén ya convencidos.

III

LA CLAVE

EN ESTE PUNTO, Rawls tuvo una idea. Me aventuraré a decir que fue una de las más bellas ideas de la historia de la teoría social y política⁴. Aunque el problema de Rawls procedía de la historia de la filosofía moral, e incluso en la década de los 50 tenía un sabor un tanto antiguo, con su elaborado interés por la doctrina más bien singular del intuicionismo, su idea fue, como dirían los jugadores de ajedrez, hipermoderna. Rawls proponía construir un modelo formal de una sociedad de individuos racionalmente egoístas, a quienes él imaginaria entregados a lo que la teoría moderna de la elección racional llama *un juego de regateo*. Su intuición consistió en que, si él constituía el juego de regateo a lo largo de las líneas sugeridas por la tradición contractual de la teoría política —es decir, si proponía un grupo de individuos cuya naturaleza y cuyos motivos fuesen los que habitualmente se suponen en la teoría del contrato—, entonces, con una sola compulsión adicional cuasi-formal, sustantivamente vacía, podría demostrar, como un teorema formal de la teoría de la elección racional, que la solución al juego del regateo era un principio moral que tenía las características de la constructividad, la coherencia con nuestras convicciones morales y la racionalidad, y que destinaba un espacio independiente a la noción de lo justo, a la vez que reconocía la dignidad y el valor de la personalidad moral. La compulsión descubierta por Rawls era tan pequeña, tan natural, era de tal evidencia que se trataba de una compulsión bajo la que se avendría a operar cualquier persona que pretendiese tener una moralidad, que Rawls, si podía demostrar su teorema, estaría en condiciones de decir a un lector:

Si es usted un agente racionalmente egoísta, y si usted pretende tener una moralidad, debe reconocer como obligatorio para usted el principio moral que yo enunciaré.

⁴ Puede parecer extraño describir una idea filosófica como "bella", pero los matemáticos están acostumbrados a aplicar términos de valoración estética a ideas abstractas, y la teoría de Rawls es, a mi juicio, una operación sencilla, elegante, formal, inserta en una enorme cantidad de ejemplificación sustantiva, y casi oscurecida por ella.

No es de extrañar que ésta sea, casi exactamente, la demanda formulada por Immanuel Kant para su Imperativo Categórico en la *Fundamentación de la metafísica de las costumbres*.

El juego del regateo de Rawls es un juego cooperativo de no-suma-cero, cuyo objeto es el de que los jugadores lleguen a un acuerdo unánime sobre un conjunto de principios que, en adelante, servirán como criterios para valorar las instituciones o las costumbres dentro de las cuales interactúan los jugadores¹. El juego consiste en una serie de propuestas sometidas por cada jugador, sucesivamente, a la consideración de todos los demás, y el juego termina cuando hay un acuerdo unánime sobre un determinado conjunto de principios. Se supone que los jugadores son racionalmente egoístas, como en todos esos juegos, pero también se supone que actúan bajo una sola compulsión adicional que no puede deducirse de la definición del egoísmo racional: se declara que, una vez que han alcanzado un acuerdo acerca de un conjunto de principios, aunque éstos hayan sido elegidos sobre la base de un cálculo egoísta, se someterán a esos principios en todos los casos futuros, incluidos aquellos en que tal sometimiento no sea conveniente para su personal interés —ni siquiera teniendo en cuenta todos los efectos secundarios y a largo plazo. Como Rawls dice en "La justicia como imparcialidad": "tener una moralidad es como haber adquirido un firme compromiso previo; porque debemos atenernos a los principios de la moralidad, aunque sea en perjuicio propio". La solución propuesta por Rawls a este juego del regateo consiste, naturalmente, en los ya famosos Dos Principios de Justicia.

Dentro de un momento, nos detendremos a examinar la primera caracterización de Rawls del juego del regateo y de su solución — lo que yo llamaré la Primera Forma del Modelo. Pero vale la pena de que, por un instante, nos paremos a reflexionar sobre la elegancia y la belleza de esta idea. Es la idea central de la filosofía de Rawls, y, a través de todas las transformaciones y elaboraciones de la teoría, persiste como la clave de su pensamiento. La verdadera fuerza de la idea puede apreciarse si recordamos algunos hechos relacionados con la historia y con el desarrollo de la teoría moral racional.

En los párrafos iniciales de la *Fundamentación de la metafísica de las costumbres*, Kant escribe que "todo conocimiento racional es, o bien

¹ Para una explicación de una parte de la terminología de la teoría del juego, y para una discusión de ciertas ambigüedades en la caracterización de Rawls del juego del regateo, ver más adelante el capítulo V y también el capítulo XV.

material y relativo a algún objeto, o bien *formal* y relativo solamente a la forma del entendimiento y de la razón en sí mismos — a las normas universales del pensamiento en cuanto tales, independientemente de las diferencias de los objetos"⁶. En sus actividades teóricas, la razón del hombre se guía por las leyes de la lógica — por la ley de contradicción, por las reglas de la inferencia silogística, etc. Estas leyes son puramente formales; tratan de la *forma* de nuestras afirmaciones (tales como "Todas las *A* son *B*", o "Si *p*, luego *q*"). Nada dicen acerca de la *materia* o contenido de las demandas del conocimiento (tales como si este tipo particular de *A* es realmente una *B*). El hecho fundamental en relación con la lógica, como todo alumno principiante aprende, consiste en que, mientras las leyes formales del razonamiento pueden *excluir* ciertas proposiciones como auto-contradictorias simplemente en virtud de su forma, y en consecuencia, como seguramente falsas, esas leyes no pueden *incluir* precisamente aquellas proposiciones cuyo contenido las hace verdaderas. "Todos los hombres son inmortales" es tan legítimo como "Todos los hombres son mortales", en lo que a la lógica se refiere.

En la ética, o en lo que Kant llamó el empleo práctico de la razón es posible encontrar leyes puramente formales a las que todos los principios morales deben ajustarse, aunque, como es de esperar, tienden a ser más discutibles que la ley de contradicción o las reglas de inferencia. Una de esas leyes es la ley de eficiencia o de prudencia, que dice, para expresarlo del modo más sencillo posible: "cualquier cosa que tú hagas, hazla al precio más bajo posible, pero debes tener en cuenta los riesgos". Una norma de este tipo tiene algún defecto, como lo tiene la ley de contradicción. Si yo quiero ir de Nueva York a San Francisco pagando lo menos posible, siempre que pueda realizar el viaje en veinticuatro horas, la ley de prudencia me dice que compre un billete de avión de clase turista, mejor que uno de primera clase. Pero, naturalmente, ni la ley de prudencia ni ninguna otra ley puramente formal de la razón práctica puede decirme si debo ir a San Francisco. Y esa es la cuestión. Cuando se trata de los fines, o de los objetivos, o de los propósitos de mi acción, parece que la razón sola es incapaz de guiarme.

Kant pensaba que el fracaso a la hora de proporcionar alguna justificación filosófica de la materia o del contenido de la razón teórica nos

⁶ Para una discusión de la *Fundamentación*, y para una defensa de la interpretación de la filosofía de Kant adoptada en este ensayo, ver mi *The Autonomy of Reason: A Commentary on Kant's Groundwork of the Metaphysics of Morals* (Harper & Row, 1973).

reduciría al estéril escepticismo de David Hume. Consideraba que un fracaso análogo en la ética nos dejaría sólo con los principios simplemente formales de eficiencia o de prudencia. No tendríamos bases racionales objetivas para elegir un sistema de fines o de propósitos en lugar de otro, y no dispondríamos, en consecuencia, de nada que se pareciera a unos principios morales sustantivos, objetivos. Así, pues, llegaba a la conclusión de que, o bien nos veríamos obligados a refugiarnos en la subjetividad de la prudencia, como hace, en última instancia, el utilitarismo, a pesar de todos sus esfuerzos en contrario, o bien nos encontraríamos en situación tan desesperada que tendríamos que proponer, simplemente, unos principios morales objetivos sustantivos, sin una sugerencia de argumentación racional, como hace el intuicionismo. En la *Fundamentación*, Kant luchó por huir de este dilema, intentando la imposible hazaña de deducir unas conclusiones sustantivas de unas premisas puramente formales.

La brillantez de la idea de Rawls consiste en su promesa de un camino de salida del atolladero al que Kant había conducido a la teoría moral, atolladero que, como hemos visto ya, seguía bloqueando a los filósofos morales casi dos siglos después. Mediante el recurso al juego del regateo, Rawls espera derivar principios sustantivos de unas premisas que, sin ser puramente formales, tampoco son manifiestamente materiales. La compulsión del compromiso es una compulsión de procedimiento, una premisa cuasi-formal que no hace referencia alguna a unos objetivos específicos a los que deban aspirar los que toman parte en el juego del regateo. La compulsión dice, sencillamente: "Una vez que hayan alcanzado un principio satisfactorio, deben estar ustedes dispuestos a observarlo, en todo momento y en cualquier circunstancia". No se establecen límites acerca del principio que deben adoptar, ni se exige de los jugadores que adopten su principio por razones "éticas" más que por motivos de interés propio.

Incluso cuando el teorema de Rawls pueda quedar establecido, el escéptico moral que atiende a su propio interés también puede negarse a adquirir un compromiso de-una-vez-para-siempre, aun cuando se trate de un principio elegido por motivos de conveniencia propia. La fidelidad al principio, en última instancia, no es deducible de la simple racionalidad formal; por lo menos, no lo es sin unos supuestos metafísicos más bien poderosos acerca del carácter intemporal del agente moral (*qua noumenon*, en el lenguaje de Kant). Pero Rawls podrá resolver la disputa entre el utilitarista y el intuicionista, pues ambos reconocen la obligatoriedad de los principios, y podrá hacer bastante más

también. Mediante esta operación, habrá alcanzando para sus principios de justicia el mismo tipo de *status* condicionalmente *a priori* que Kant reivindicaba para el sistema de principios del puro entendimiento en la *Crítica de la razón pura*⁷.

SEGUNDA PARTE EL DESARROLLO DE LA TEORÍA

⁷ Para una amplia discusión de la relación entre Rawls y Kant, ver tercera parte. Ver también, *The Autonomy of Reason*, Conclusión.

SEGUNDA PARTE
EL DESARROLLO DE LA TEORÍA

IV

LA PRIMERA FORMA DEL MODELO

EN SU PRIMERA forma, el modelo de Rawls es sencillo, claro, elegante, y —como veremos— puede ser blanco de objeciones devastadoras. A pesar de sus deficiencias, sin embargo, estoy dispuesto a sostener que la primera forma del modelo es el verdadero fundamento sobre el que se construye todo el resto de la teoría de Rawls. En consecuencia, lo someteré a un detenido examen⁸.

Rawls declara que considera la justicia "como una virtud de las instituciones sociales, o de lo que yo llamaré prácticas (*practices*)". Una práctica se define como "toda forma de actividad especificada por un sistema de normas que definen oficios, funciones, movimientos, castigos, defensas, etc., y que dan su estructura a la actividad". La noción de una práctica es fundamental para el modelo, porque el juego del regateo es un juego cuyo resultado o solución es un acuerdo sobre los principios que han de emplearse para determinar si las prácticas son aceptables — si son "justas" en su estructura más que en su simple utilización. En esta concepción inicial de una práctica, están claras varias cosas que después se tornan más bien confusas. La primera es que cada práctica debe ser definida por sus normas y sus funciones, independientemente de las recompensas o de las cargas asignadas a las diversas funciones. En caso contrario, sería imposible exigir de una práctica dada, como Rawls hace dos páginas después, que "el hombre representativo en todo oficio o posición definidos por tal práctica, cuando él la considera ya como una empresa en marcha, deba encontrar razonable preferir su situación y sus perspectivas con la desigualdad, a lo

⁸ En este capítulo y en el siguiente, me referiré, sobre todo, al ensayo de 1958, "Justice as Fairness", siguiendo la reimpression de Laslett y Runciman. Parte de lo que digo acerca de esta versión de la teoría de Rawls está sacado de mi artículo "A Refutation of Rawls's Theorem on Justice", *J. Phil.*, 63 (31 de marzo de 1966) págs. 179-190.

⁹ "La Justicia como imparcialidad". Rawls convirtió más tarde este ensayo en el capítulo I de su *A Theory of Justice* (*Teoría de la Justicia*, Fondo de Cultura Económica, 1979) [N. del E.]

que serían en el marco de la práctica [probablemente, la *misma* práctica] sin ella". En efecto, en cuanto a variaciones marginales en los sueldos, esta concepción de la relación entre una práctica y su patrón de recompensas es aceptable. Una práctica o institución como, por ejemplo, un hospital es, indudablemente, la *misma* práctica o institución, si se introduce un ligero cambio en las remuneraciones, de modo que se reduzca la diferencia entre los jefes de servicio y los internos, o entre la enfermera-jefe y las enfermeras de sala. Pero si se piensa en redistribuciones más importantes, como la que consistiría, por ejemplo, en que el jefe de cirujanos viera su elevado sueldo reducido al nivel del sueldo de un practicante, mientras los obreros de sostenimiento pasaban a cobrar más que los médicos del equipo de dirección, no está claro, en absoluto, que consideremos ni podamos considerar las dos situaciones como ejemplos de la *misma* práctica con diferentes sueldos. En el primer caso, los sueldos asignados a una función son, en cierto modo, una parte de esa función; en el segundo caso, importantes factores secundarios como las relaciones de poder y de autoridad, esquemas de respeto y formas de coordinación funcional, se verían alterados por algo más que por marginales reajustes de sueldos. Más adelante, Rawls traslada su atención desde esas prácticas de pequeña escala, que los economistas conocen como el micro-nivel, hacia el sistema fundamental de las instituciones sociales y económicas de una sociedad como conjunto, correspondiente al macro-nivel del análisis en economía. Pero esto plantea un nuevo problema, que había sido sutilmente tratado por la anterior y más simple concepción de una práctica.

La teoría de Rawls habla de las expectativas de los hombres *representativos*. En el primer modelo, cada hombre representativo debe creer que una distribución desigual actúa en favor suyo. En el segundo modelo y en los sucesivos, se introduce el llamado "principio de la diferencia", según el cual el hombre representativo menos favorecido debe esperar que las desigualdades actúen en su beneficio. Pero, en uno y otro caso, se supone que podemos identificar a los diferentes hombres representativos. Ahora bien: en la concepción original de las prácticas, los hombres representativos son claramente identificados por las reglas de la práctica, de una manera análoga, o tal vez incluso idéntica, a la forma en que los jugadores son identificados por las reglas de un juego. En el juego del béisbol, por ejemplo, está claro, según las reglas, que hay nueve jugadores en cada equipo (sin contar los *pinch hitters*, los *relief pitchers*, los *managers*, los *coaches*, etc.). Si surgiese una discusión acerca de la imparcialidad del juego del béisbol, no se daría crédi-

to alguno a una falsa argumentación destinada a demostrar que el juego satisface el principio de diferencia de Rawls, mientras interpretamos las reglas en el sentido de que definen cuatro "hombres representativos", concretamente el *outfielder*, el *infielder*, el *pitcher* y el *catcher*. Pero, cuando preguntamos qué individuos de una sociedad industrial son los que hemos de considerar que forman el "grupo menos favorecido", no es evidente, en absoluto, que tengamos que mirar a la clase obrera como conjunto, o a los obreros no calificados, o a las mujeres negras no calificadas, o a cualquier otro sector. La extensión más natural de la noción de una práctica de Rawls a una sociedad entera sería un análisis marxiano funcionalista, en términos de la estructura de producción y de las relaciones sociales de producción, pero Rawls, por su parte, tiende a hablar el lenguaje de la sociología americana liberal, que sustituye el concepto de clase económica con una escala de *status* socioeconómicos. Una vez dado este paso, los puntos aislados se hacen más convencionales que objetivos, y no es extraño encontrar a Rawls flirteando con la idea de una conexión en cadena como un medio de soslayar esas dificultades.

Una vez esbozada la noción de una práctica, Rawls desarrolla ahora los elementos de su juego del regateo. Tenemos que imaginar una sociedad de agentes racionales, entre los que se halla establecido ya un sistema de prácticas. Los miembros de la sociedad tienen las siguientes características:

1. Adoptan decisiones sobre la base de un interés propio inteligente, y son capaces de descubrir sus preferencias personales y de valorar, con un razonable acierto, las consecuencias de sus acciones y también de las ajenas⁹.
2. Tienen necesidades e intereses aproximadamente semejantes, o, por lo menos, necesidades e intereses que hacen que sea racional una interesada cooperación entre ellas.
3. Son "suficientemente iguales en poder y en habilidad para ga-

⁹ Como Rawls pretende contar, en este punto, con el tipo de unanimidad cuasi-ordenadora habitualmente asociada al nombre de Pareto, se limita a suponer una preferencia ordinal. Sin embargo, a la luz de los cálculos de "expectativa razonable" que sus hombres representativos tienen que hacer, parece que será necesaria una cierta concepción de utilidad cardinal. En la versión final del modelo, cuando se ha descorrido el velo de la ignorancia, Rawls arguye que los participantes en el juego del regateo verán que es racional adoptar la regla maximin de elección en condiciones de incertidumbre. Sus razones para esta pretensión, como veremos, presuponen, no simplemente funciones de utilidad cardinal, sino también funciones de utilidad con un punto-cero natural. Ver más adelante, cuarta parte.

rantizar que, en circunstancias normales, ninguno puede dominar a los otros".

4. No son envidiosos; esto quiere decir que "el simple conocimiento o percepción de la diferencia entre su situación y la de los otros no es, dentro de ciertos límites y por sí solo, motivo de grave insatisfacción".

La condición 1 es el supuesto clásico de un egoísmo racional, con el que comienzan la economía del bienestar y la teoría del juego. La condición 2 garantiza que los individuos seguirán comprometidos con las prácticas, porque, si la cooperación es contraria a los intereses de un número importante de individuos la sociedad se desintegrará, o bien en la guerra de todos contra todos, o bien en una simple multiplicidad atomizada de unidades independientes. La condición 3 garantiza que, en contextos en que la diferenciación y la integración funcionales sean mutuamente beneficiosas, el egoísmo racional de todos encontrará más conveniente algún modo de regateo y de cambio, que una simple dominación y sumisión. (La condición 2 pudo prevalecer, incluso, en una situación de esclavitud, en la que una parte de la sociedad, sólo por la fuerza, fue capaz de dominar al resto, al que arrancó así unos términos sumamente desfavorables de "cooperación", que fueron, de todos modos, superiores a los de una completa ruptura de todas las prácticas sociales.)

La condición 4, la estipulación de "no-envidia", ha dado origen a muchos comentarios en la literatura sobre Rawls. La razón de su presencia, como veremos, es la de que, sin ella, no se puede utilizar convenientemente la relación cuasi-ordenadora de la preferencia unánime introducida por Pareto. Si se supone que los individuos tienen una positiva utilidad marginal por cada artículo o mercancía producido, o por determinado índice de artículo, o por dinero (cualquiera que sea la medida que Rawls decida utilizar al discutir las distribuciones por las funciones definidas por una práctica); y si cada individuo prefiere una distribución a otra, cuando —y sólo cuando— la primera le da más utilidad de cada artículo o mercancía, o un índice más alto de artículos, o más dinero que la segunda (suponiendo que cualesquiera otras necesidades y condiciones sean satisfechas por ambas distribuciones); en resumen, si en este especialísimo y sumamente técnico sentido del término, los participantes en el juego del regateo son "no-envidiosos"; entonces, se ofrece la posibilidad, por razones que luego se explicarán, de que una distribución desigual sea unánimemente preferida por los jugadores a una distribución igual pero de nivel más bajo.

Después, en *Teoría de la Justicia*, Rawls se entrega a cierta psicología moral especulativa, extremadamente elaborada, en parte, al menos, para proporcionar una base factual al supuesto de la no-envidia. Pero, si no me equivoco, esas especulaciones son rigurosamente *post hoc*. La verdadera razón del supuesto de la no-envidia es puramente técnica, y guarda relación con los supuestos exigidos por los modos de razonamiento cuasi-económico que Rawls gusta de emplear.

La clave para la solución del problema del regateo, en opinión de Rawls, es un hecho fundamental de la vida económica — lo que yo llamaré la posibilidad de un *excedente de la desigualdad*. Ocurre, a veces, que una distribución desigual de sueldos para las funciones de una práctica da como resultado o hace posible una situación que es unánimemente preferida a la de esa misma práctica sin la desigualdad. Todo lo que se requiere es que la desigualdad, al originar un esfuerzo mayor o un nivel más alto de habilidad o cualquier otra mejora, incremente el producto de la práctica lo suficiente para que, una vez que las funciones menos pagadas se paguen al nivel previo de la distribución igual, y todas las demás, las funciones mejor pagadas, se paguen a niveles suficientes para aportar el producto incrementado, quede algún excedente con el que pueda elevarse, al menos marginalmente, la paga de las funciones peor remuneradas.

Observamos, por ejemplo, un taller en el que se fabrican zapatos por medio de una práctica productiva en la que se definen seis funciones. Supongamos que hay esa diferenciación funcional, pero no una variación en la escala de salarios. Sesenta obreros ocupan las seis funciones (no necesariamente diez para cada función, desde luego), y el ingreso anual neto del taller, antes del pago de salarios, es de \$600 000. (Haremos caso omiso del beneficio, porque, hablando en general, Rawls también lo hace). Cada obrero cobra \$10 000 anuales.

En esta situación, los obreros trabajan a un ritmo más bien lento. Pero podemos imaginar que, si se introdujese un ritmo de trabajo más rápido en dos de las funciones, el producto del taller aumentaría considerablemente. Esto se debe a que las características técnicas del proceso productivo crean un atasco en esos dos puntos, de tal modo que toda la empresa sería mucho más eficiente si pudiera inducirse a esos obreros a que trabajasen más de prisa. Supongamos que hay cuatro obreros en una de las tareas del atasco y seis en la otra. Supongamos también que, para compensar a esos obreros por la molestia y por el mayor esfuerzo del trabajo acelerado —a fin de inducirlos a que se ha-

gan cargo de esas tareas—, tenemos que pagar a cada uno de ellos \$15 000, en lugar de \$10 000. Como se necesita un total de diez obreros para las tareas del atasco, tendremos que encontrar \$50 000 más para poner el taller a este ritmo más productivo.

Ahora bien: hay tres posibilidades. El ingreso neto del taller con el nuevo sistema, que por hipótesis será superior al original de \$600 000 puede ser menor que \$650 000, exactamente igual a \$650 000, o superior a \$650 000. Si es inferior a \$650 000, tendremos que *reducir* los salarios de algunos de los otros cincuenta obreros, a fin de disponer de los \$5 000 más que se necesitan para cada trabajador del nuevo sistema. El taller, como conjunto, será más productivo, pero nunca podremos conseguir que los sesenta obreros, en su totalidad, estén de acuerdo con el cambio, suponiendo que sean racionalmente egoístas.

Si el ingreso neto es exactamente igual a \$650 000, los otros cincuenta obreros serán indiferentes entre el nuevo sistema y el antiguo, porque sus tareas y sus salarios no cambiarán. Los diez obreros rápidos preferirán, por hipótesis, el nuevo sistema al antiguo, porque se ha supuesto que sus salarios de \$15 000 eran suficientes para inducirlos a unas tareas más duras, lo que sólo puede significar que *prefieren* el nuevo trabajo pagado con el nuevo salario, a su trabajo antiguo pagado con el antiguo salario.

Pero si el ingreso neto *supera* los \$650 000, podrán darse a los obreros que trabajan más duramente sus salarios incrementados, y todavía quedará algo para subir los salarios de los cincuenta obreros restantes por encima del nivel básico anterior de salario igual (bajo). Por ejemplo, si el ingreso neto asciende a \$700 000, entonces, una vez que se hayan pagado sus \$10 000 a cada uno de los cincuenta obreros normales y sus \$15 000 a cada uno de los obreros rápidos, habrá quedado un depósito de \$50 000, que puede distribuirse entre los cincuenta obreros normales, elevando sus salarios a \$11 000 cada uno. *Esos \$50 000 son un excedente de la desigualdad* — son el ingreso excedente que resta después de haber pagado a todos los ocupantes de las funciones de una práctica desigualmente remunerada lo suficiente para inducirlos a que realicen las diversas funciones.

Y ahora quedará clara la cuestión del supuesto de la no-envidia. Si hubiéramos de permitir que los juicios de los participantes en nuestro juego del regateo se viesan influidos por la "envidia", un jugador podría mostrarse tan resentido por el hecho de que su compañero ganase un salario de \$15 000, que preferiría seguir con el sistema inicial

de pagas iguales, *aunque para ello tuviera que renunciar a una subida salarial de \$1 000*. Pero, si excluimos la envidia, podemos estar seguros de que, cuando un cálculo objetivo revele que una determinada distribución de salarios en una práctica producirá un excedente de desigualdad, deberemos llegar a la conclusión de que habrá una posible distribución de ese excedente, en virtud de la cual la práctica *con* la desigualdad será unánimemente preferible a la misma práctica sin la desigualdad¹⁰.

Supongamos ahora que una sociedad de individuos que se ajusta a esta descripción se halla empeñada en sus prácticas establecidas, y que, de cuando en cuando, los individuos formulan quejas contra esas prácticas. Cada queja adopta la forma de una reivindicación en el sentido de que una determinada función en una práctica debería recibir mayores remuneraciones o llevar implícitos más brillantes reconocimientos, o de que las normas que rigen la asignación de unos individuos a unas funciones deben modificarse, o tal vez incluso la de que una cierta práctica nueva debe sustituir a una práctica existente. (Rawls no se preocupa de las quejas según las cuales las normas indudablemente correctas de una práctica están siendo aplicadas incorrectamente. Es decir, no se interesa por la simple justicia procesal.) Supongamos también que, antes de intentar resolver determinadas reivindicaciones particulares, la sociedad decide elegir, de una vez para siempre, los principios generales por los que han de resolverse todas las disputas futuras. Rawls sugiere que imaginemos que los miembros de la sociedad intervienen, pues, en el juego del regateo descrito más arriba. Según Rawls, cada jugador razonará así: "Yo quiero todo lo que pueda alcanzar. Por consiguiente, defenderé un conjunto de principios que convenga a

¹⁰ Si tuviéramos que intentar formalizar esta discusión matemáticamente, surgirían ciertos problemas un tanto delicados relacionados con intervalos continuos carentes de límites inferiores. Podríamos abordar esos problemas suponiendo que los cambios salariales sólo eran posibles en ligeros incrementos, como de \$ 100 anuales, sobre todo si hemos supuesto también que las funciones de utilidad de los obreros eran insensibles a cambios salariales inferiores a esa cantidad. Desgraciadamente, en ese caso, tendríamos que enfrentarnos con el hecho de que la indiferencia no sería transitiva. Más aún: podría haber excedentes de desigualdad demasiado pequeños para ser distribuidos entre todos los obreros de una práctica en cantidades perceptibles. Sin embargo, creo que podemos excusar razonablemente a Rawls de preocuparse por estas cuestiones. Como decía Aristóteles, "nuestra discusión será adecuada, si tiene tanta claridad como la materia lo permita, porque la exactitud no debe perseguirse en todas las discusiones, como no se persigue en todos los productos de las artes" (*Moral a Nicómaco*, I, 3).

mis circunstancias, aunque mejor sería que no me apresurase a proponer principios favorables al aprieto en que me encuentro ahora, porque me comprometo para todo el futuro, cualquiera que éste pueda ser. Pero mis adversarios en este juego no son tontos, y rechazarán, naturalmente, esas propuestas indirectas, si no les ofrecen, en realidad, un marco más ajustado a sus propias preferencias. Está claro que, entonces, yo debo insistir en la igualdad. Ellos no me darán más, y yo no me conformaré con menos. Pero detengámonos un momento. Supongamos que cierta distribución desigual pueda incrementar el producto de nuestras prácticas hasta el punto de que origine un excedente de desigualdad. En este caso, si yo propongo una redistribución de ese excedente que beneficie a todos los hombres representativos, puedo estar seguro de que saldré absolutamente más favorecido que con un patrón de igualdad. Como yo no me preocupo del aumento de la ganancia de mis compañeros, siempre que yo me beneficie también, permitiré cuantas desigualdades actúen en beneficio de todos. Pero *no* aceptaré ninguna distribución desigual que empuje a determinadas funciones por debajo de la línea básica de la igualdad, para elevar por encima de ella a otras. No estoy dispuesto a correr el riesgo de quedarme en esa función rebajada". Y, en consecuencia, los jugadores se ponen de acuerdo sobre estos dos principios:

Primero: toda persona que participe en una práctica, o sea afectada por ella, tiene un derecho igual a la más amplia libertad compatible con una libertad igual para todos. Y segundo: las desigualdades son arbitrarias, a menos que sea razonable esperar que actuarán en beneficio de todos, y que las posiciones y oficios a los que se hallan unidas, o desde los que pueden ser conquistadas, están abiertos a todos.

V

UNA CRÍTICA DE LA PRIMERA FORMA DEL MODELO

ANTES DE que nos hallemos en condiciones de valorar la pretensión de Rawls de que nos ha proporcionado un "esbozo de una demostración", es preciso contestar a cuatro preguntas, por lo menos, relacionadas con el juego del regateo y con esos principios. Primera: ¿cuáles son las condiciones de fondo o los supuestos implícitos del juego del regateo? Segunda: ¿cómo debemos interpretar, exactamente, los dos principios? Tercera: ¿proporcionan los dos principios, clara e inequívocamente, una valoración o catalogación de las prácticas y unos esquemas alternativos de distribución dentro de esas prácticas? Y por último: ¿los individuos situados como Rawls supone, empeñados en un juego de regateo, se ponen realmente de acuerdo sobre los dos principios por decisión unánime? En resumen: ¿son los dos principios, realmente, la solución al juego del regateo? Dedicaré un tiempo considerable a responder a estas preguntas, a pesar de que están dirigidas a una primitiva y ya descartada versión de la teoría de Rawls, porque, a mi juicio, la mayor parte del sucesivo desarrollo de la teoría puede seguirse a través de las dificultades planteadas por estas preguntas relativas a la primera forma del modelo.

LOS SUPUESTOS DE FONDO

Respecto a los supuestos de fondo o precondiciones del juego del regateo, deben señalarse diversos puntos. Las características de los jugadores y de su situación, citadas más arriba, con la excepción de la cuarta exigencia, la no-envidia, constituyen lo que Rawls denomina las "condiciones de la justicia". Definen las circunstancias en que surgirían, de un modo natural, las cuestiones de la justicia distributiva. Siguiendo una tradición que se remonta hasta Hume y aun más allá, Rawls supone que en un mundo de superabundancia, o en ausencia de toda común necesidad de cooperación, o entre seres demasiado agresivos para disciplinarse ni siquiera en su propio beneficio, o demasiado cordiales y altruistas para estar dispuestos a defender sus propios derechos contra

los ajenos, o, finalmente, en un mundo en el que unas personas tienen toda la capacidad necesaria para dominar a los demás, así como la voluntad de hacerlo, los debates acerca de la justicia serían una farsa. En un mundo semejante, como diría Locke, o no se plantearía la cuestión, o adoptaría, necesariamente, la forma de una "apelación al cielo".

Todo esto es bien sabido e indiscutible. Pero del carácter del propio juego del regateo se derivan consecuencias algo más importantes. Ante todo, el juego es una confrontación de regateo cara-a-cara, en la que cada propuesta es formulada públicamente y es oída claramente por todos los participantes. Este rasgo de la publicidad es un elemento estipulado del juego, y no una observación sociológica, política o psicológica en torno a la conducta del regateo. Los participantes son también formalmente iguales en la facultad de adoptar decisiones, toda vez que el juego requiere unanimidad. Se supone, sencillamente, como parte de la estructura del juego, que ningún jugador utilizará amenazas ni coacciones para arrancar a los otros un acuerdo sobre unos principios. Como en el juego no se establecen límites de tiempo, se trata de un juego simétrico en el que ningún jugador obtiene una ventaja, por ejemplo, por salir el primero, o, lo que es aun más importante, por ser el último en salir¹¹. Todos los jugadores tienen un derecho igual a participar en la discusión de propuestas alternativas, un derecho que se deriva, directamente, de su facultad de detener el regateo mediante la negativa a dar su aprobación hasta que haya sido escuchado.

Estos distintos derechos, generalmente reunidos bajo el epígrafe de "libertad política", se hallan establecidos, en este primer modelo, dentro del juego mismo; no constituyen objeto de deliberación dentro del juego. De ahí que tengan "prioridad", en el sentido de que son presupuestos de los procedimientos del juego. Esta relación entre elementos formales o de procedimiento del juego y las propuestas sustantivas de principios de distribución es un modelo perfecto del tradicional análisis liberal de la relación entre las garantías formales políticas y legales del estado burgués liberal y las disposiciones económicas o patrones de distribución a los que se llega por la acción del mercado libre. En la nueva economía del bienestar, se supone que las libertades puramente políticas pueden establecerse mediante recursos constitucionales o políticos, y que, dentro de ese marco de "libertad igual", pueden debatirse varias políticas distributivas o redistributivas. En la primera versión

¹¹ Para una discusión del tema de la simetría, ver más adelante en este capítulo.

de su modelo, Rawls establece esa concepción tradicional de la relación entre libertad política y distribución económica dentro de la estructura del juego del regateo. En modelos ulteriores, como veremos, desvía significativamente su base, haciendo de la libertad política uno de los principios elegidos por los que participan en el juego. Este cambio, según hemos de ver, crea a Rawls enormes problemas lógicos y conceptuales.

LA INTERPRETACIÓN DE LOS DOS PRINCIPIOS

¿Cómo hemos de interpretar los dos principios? Por razones de comodidad, llamaremos Principio I a la primera cláusula, relativa a "la más amplia libertad compatible con una libertad igual para todos". La exigencia de que las desigualdades actúen en beneficio de todos será designada como Principio IIa, y la condición de que las posiciones favorcidas estén abiertas a todos se reconocerá como Principio IIb. Cada uno de estos componentes plantea algunos problemas de interpretación.

El Principio I es desconcertante porque se refiere a la "libertad" más que a la riqueza, o a los ingresos, o a las remuneraciones. La idea central que hay detrás de los principios de Rawls parece bastante clara: el producto o los ingresos de una práctica han de distribuirse por igual, a menos que pueda hacerse actuar en beneficio de todos algún patrón de distribución desigual, del modo esbozado más arriba, y que todos tengan un acceso posible a las funciones mejor pagadas. A fin de garantizar una simple eficacia —garantizar que la práctica esté cerca de su "frontera de producción"—, Rawls estipula la más amplia libertad compatible con una libertad igual para todos. De otro modo, el principio no podría hacernos avanzar desde una distribución igual de un bajo nivel de paga, en la que sólo se consumía una cierta porción del producto total de la práctica, hasta una situación de paga igual más alta, en la que se distribuía todo el producto. Como lo segundo sería unánimemente preferido a lo primero, los jugadores racionalmente egoístas insistirían en que se alcanzase siempre que fuese posible.

Pero, si es esto lo que Rawls quiere decir, ¿por qué habla de "libertad igual"? A la luz de la subsiguiente revisión de sus dos principios y de la introducción de la norma de la prioridad de la libertad, sería natural que aquí le interpretásemos en el sentido de que pretende distinguir entre cuestiones de libertad política y cuestiones de distribución económica. La confusión podría reducirse, entonces, a un simple error

en la formulación. Pero hay dos razones para rechazar esta explicación, y ambas, a mi juicio, decisivas. En primer lugar, cuando el propio Rawls glosa el segundo principio, aclara que lo propone como una matización del primer principio de igualdad, más que como un principio distinto orientado a un tema diferente. Dice: "El segundo principio define qué clases de desigualdades son admisibles; especifica cómo puede desecharse el supuesto establecido por el primer principio". Adviértase que Rawls no dice que el segundo principio especifique las condiciones en que puede prescindirse de la libertad política a cambio del beneficio económico. En otras palabras, no presenta el segundo principio como base para anular el primero. Más bien presenta el segundo principio como si estableciese las bases sobre las cuales puede desecharse el supuesto (de distribución igual). Lo que está en juego, evidentemente, es la cuestión de cuándo una distribución desigual de salarios puede sustituir justamente a una distribución igual de salarios, y no la cuestión, totalmente distinta, de cuándo un cierto patrón de remuneraciones de un determinado tipo de bienes (riquezas, etc.) puede ser invocado como justificación para desviarse de una distribución igual de un diferente tipo de bienes (concretamente, la libertad). Esta interpretación se ve reforzada por el hecho de que la noción misma de "libertad igual" en un sentido *distributivo* es sumamente confusa. Decir de dos personas que son "iguales ante la ley" es decir algo muy preciso acerca del modo en que son tratadas por la ley. Decir que todas las personas serán iguales ante la ley es, en el contexto de la teoría política de finales del siglo XVIII y comienzos del XIX, decir algo importante y significativo acerca de los ordenamientos legal y político de una sociedad. Pero decir que todas las personas recibirán "la más amplia libertad compatible con una libertad igual para todos" es, para decirlo suavemente, misterioso. A no ser que invoquemos una cierta noción sumamente metafórica de un "espacio moral de derechos individuales" y hagamos ciertas insinuaciones pseudo-formales de un género cuasi-topológico en torno a la noción de compactas esferas de derechos individuales que se expanden y se deforman continuamente hasta que llenan todo el espacio social, o cualquier otro, la noción de la "más amplia libertad (política) compatible con una libertad igual para todos" no puede aclararse, sencillamente. Pero podemos aclarar mucho, ciertamente, la noción de una muy amplia distribución igual de bienes.

La segunda consideración que apoya mi interpretación del primer principio es el hecho de que, gracias a ella, el "teorema" de Rawls pa-

rece aceptable, al menos inicialmente. La demostración del teorema implicará, sencillamente, una invocación a la concepción de la optimalidad de Pareto, en el entendimiento de que la cuasi-ordenación de las distribuciones alternativas ha de hacerse con respecto a la situación original o de línea de partida de la distribución igual. Tal línea de razonamiento en apoyo de los dos principios sólo tendrá sentido si el primer principio se interpreta como una norma *prima facie* de la distribución igual, y si el segundo principio se interpreta como una excusa para las desviaciones de la igualdad distributiva.

En relación con el Principio IIa, baste señalar que Rawls invoca aquí lo que a veces se llama preferencia fuerte de Pareto, más que preferencia débil de Pareto. Las desigualdades no pueden justificarse simplemente sobre la base de que no perjudican a nadie y de que benefician, por lo menos, a una persona. Más bien, sólo se justificarán si benefician a *todos*. Esta es una fuerte exigencia, y no está claro que Rawls trate, realmente, de insistir en ella. Supongamos, para volver a nuestro ejemplo del taller de zapatos, que el paso al nuevo sistema incrementa el ingreso neto del taller exactamente hasta \$650 000 al año, de tal modo que, una vez que los cincuenta obreros no afectados reciben \$10 000 cada uno y los diez obreros rápidos reciben sus \$15 000 cada uno, no queda ningún excedente de desigualdad para aliviar el trato. Los cincuenta obreros normales son indiferentes ante las dos prácticas, porque sus obligaciones y sus recompensas son las mismas en ambas. Por hipótesis, los diez obreros rápidos prefieren el nuevo ordenamiento, bajo el cual realizan un trabajo más duro y ganan más. Parece que debería ser colectivamente racional que los que participan en el juego del regateo clasificasen ese ordenamiento desigual por encima de la línea básica de la perfecta igualdad. De todos modos, a excepción de la pura envidia ante la mejora experimentada por los diez, ¿qué es lo que podría impulsar a los cincuenta a votar contra esa clasificación? Y como cada jugador reconocerá que puede haber situaciones en las que él sea uno de los beneficiarios de la desigualdad, deseará establecer esa clasificación como un principio general¹².

¹² La economía del bienestar trata los problemas de la distribución de un modo completamente independiente de las exigencias impuestas por las técnicas de la producción. Haciendo algunos convincentes supuestos de continuidad, se demuestra, pues, fácilmente, que la preferencia fuerte y la débil de Pareto son equivalentes. Cuando un solo individuo prefiere la situación A a la situación B, porque B le da una mayor cantidad de lo que se está distribuyendo, podemos, simplemente, quitarle

Con el tiempo, Rawls optó por otra interpretación del Principio IIa, según la cual la situación de los menos beneficiados debe ser maximizada (y, en consecuencia, maximizada, en orden lexicográfico, cada situación de posición superior). Este cambio tuvo el efecto secundario de promover la discusión entre la preferencia débil y la fuerte de Pareto.

El Principio IIb también es susceptible de diversas interpretaciones. "Abiertos a todos" puede significar, entre otras cosas, abiertos a todos en una justa competición, o abiertos a todos en alguna forma de rotación, o abiertos a todos mediante una asignación por sorteo. Relativamente pronto, después de la aparición de "La justicia como imparcialidad", Rawls había optado por la noción de la justa competición, pero, como veremos en seguida, siguen en pie ciertos problemas en relación con el Principio IIb.

LA APLICACIÓN DE LOS PRINCIPIOS

¿Nos facilitan los Principios I y II una norma clara e inequívoca que nos permita valorar las prácticas reales o las propuestas? Huelga decir que la cuestión *no* está en saber si nos facilitan una norma correcta de prácticas justas. Se supone que el carácter correcto de la norma se halla establecido, de un modo o de otro, por el hecho de que los dos principios son la solución del juego del regateo que Rawls ha definido. Pero antes de que podamos considerar siquiera la plausibilidad de tal pretensión, debemos preguntar si los principios nos proporcionan, efectivamente, una clasificación de las prácticas.

Supongamos que una sociedad ha adoptado los dos principios: ¿cómo se aplicarán cuando surja una disputa? Hay tres tipos de objeciones que un miembro de la sociedad podría formular contra una de sus prácticas, correspondientes a las condiciones expuestas en el Principio I, en el Principio IIa y en el Principio IIb. A mi parecer, la invocación del Principio I no plantea problemas (suponiendo, como supondré, que es un principio de distribución igual y no un principio cuasi-político de libertad igual). El Principio IIb plantea ciertos problemas un

la cuota extra, dividirla tantas veces como sea necesario, y repartirla de tal modo que todos obtengan, por lo menos, una pequeña porción más. Pero si el salario más bajo por el que los obreros harán el trabajo a un ritmo más rápido es el de \$ 15 000, al rebajar ese salario todavía un poco más, se forzará al taller a regresar a las anteriores condiciones de producción. Este es un pequeño detalle, pero tiene amplias ramificaciones, algunas de las cuales se examinan más detenidamente en la última sección de este ensayo.

tanto delicados que, de momento, voy a aplazar. Examinemos, pues, el Principio IIa, que es el meollo de la concepción de la justicia de Rawls.

¿Qué significa decir que las desigualdades actúan o no actúan en beneficio de todos? Es probable que sólo signifique que todos pueden esperar, razonablemente, que estarán mejor *con* la desigualdad que sin ella. Pero, ¿mejor que qué? Rawls dice que mejor que en la misma práctica sin la desigualdad. En resumen, la práctica con la desigualdad es unánimemente preferida a la práctica sin la desigualdad. Pero esto plantea tantos problemas como los que resuelve. Supongamos que la práctica de que se trata es una fábrica en la que unos obreros son capataces y otros atienden a las máquinas, y en la que los capataces ganan más dinero y hacen un trabajo menos duro (porque la labor de dirección es menos pesada que la labor de atender a las máquinas). ¿Cuál sería la misma práctica sin las desigualdades? Podríamos igualar la paga, naturalmente, pero no podríamos eliminar las diferentes cargas de las distintas funciones, sin cambiar la división del propio trabajo, es decir, sin cambiar la práctica.

El problema se complica cuando tenemos que comparar dos prácticas que sirven al mismo objetivo, pero que definen diferente número de funciones y distribuyen los individuos entre ellas en diferentes cantidades. Por ejemplo, ¿cómo compararemos un modo de producción de una fábrica en el que un sistema de funciones altamente estratificado define diez posiciones distintas, teniendo cada una de éstas sus funciones y pagas características, con un segundo modo de producción (destinado a producir la misma mercancía) en el que un agrupamiento más simple y más integrado de funciones origina sólo tres posiciones, desigualmente pagadas y funcionalmente diferenciadas? Cuando se nos pregunta si cada hombre representativo puede esperar, razonablemente, verse beneficiado por la segunda práctica en relación con la primera, ¿cuáles son las funciones que hemos de comparar con cuáles? Al desplazarnos, con la imaginación, de una práctica a otra, ¿qué recurso nos permite elegir en cada práctica posible el mismo hombre representativo, de modo que podamos preguntar si en cualquiera de las prácticas imaginadas ese hombre puede esperar, razonablemente, que estará mejor que en determinada práctica real que se examine?

Si recordamos que Rawls está empleando la noción de preferencia de Pareto, resultará evidente que eso le plantea todos los problemas que son tan conocidos en la economía del bienestar. "Unánimemente preferida a", tanto en su sentido débil como en el fuerte, sólo define

una cuasi-ordenación. Dos prácticas, P' y P'' , pueden ser unánimemente preferidas a una tercera práctica, P , aunque ni P' ni P'' sean unánimemente preferidas la una a la otra. Por ejemplo, el capitalismo puede ser unánimemente preferido a lo que Engels llamaba comunismo primitivo (suponiendo que pudiéramos identificar a los adecuados "hombres representativos" dentro del capitalismo, y suponiendo también que el hombre primitivo típico esté peor que el peor hombre representativo dentro del capitalismo —lo que recientes trabajos antropológicos niegan—); y el feudalismo puede también ser unánimemente preferido al comunismo primitivo (dejando a un lado violaciones del Principio IIb); pero el capitalismo *no* puede ser unánimemente preferido al feudalismo, porque podría ser, como Marx pensaba (aunque la investigación reciente tiende a negarlo), que los campesinos que abandonan los señoríos feudales para irse a los pueblos y a las ciudades sufran un descenso absoluto en su nivel de bienestar. Si tuviéramos que emplear los principios de Rawls, nos veríamos obligados a concluir que el capitalismo era justo en relación con el comunismo primitivo e injusto en relación con el feudalismo, lo cual es, por lo menos, una extraña manera de emplear la noción de justicia social. El origen del problema, gramaticalmente hablando, consiste en que la frase "en beneficio de todos" suena como una frase positiva, pero es, en realidad, una frase comparativa (es decir, suena como "bueno", pero funciona como "mejor").

Un problema análogo surge cuando alguien se decide a proponer una alteración en una práctica ya existente. Si se produce un excedente de desigualdad, habrá algún medio de distribuirlo de modo que cada posición representativa esté mejor en la nueva práctica propuesta que la posición correspondiente en la práctica existente (suponiendo, una vez más, que podamos resolver el enojoso problema de equiparar las correspondientes funciones en las diferentes prácticas). En consecuencia, la práctica propuesta será unánimemente preferida a la práctica existente (y, en consecuencia, ¡*más justa*!; otra vez, un extraño empleo). Pero, en general, habrá muchos modos diferentes de llevar a cabo esa distribución, cada uno de los cuales es unánimemente preferido a la práctica original y ninguno de los cuales sea, acaso, unánimemente preferido a cualquier otro (porque, dado un excedente de desigualdad, toda alteración en su distribución consistirá en sacar un poco del excedente de un hombre representativo para dárselo a otro).

Consideremos, por un momento, uno de los casos más simples que podrían corresponder al Principio IIa de Rawls. Supongamos una

práctica, P , que define m funciones, de modo que el número de personas que ocupan la función i es n_i , y la remuneración o paga (se supone, por comodidad, que es dinero) asignada a la función i es P_i . En consecuencia, el ingreso neto o paquete de riqueza que se divide entre los participantes en la práctica viene dado por la expresión:

$$\text{Remuneración a } P = \sum_{i=1}^m n_i P_i$$

Supongamos que todas las P_i son iguales, excepto una sola remuneración, P_j , que es mayor, y que el producto total de la práctica con esta desigualdad es mayor que sin ella. Hay dos posibilidades. La primera posibilidad es la de que no haya excedente de desigualdad. Los individuos adscritos a la función j por la promesa de una remuneración mayor, P_j , incrementan el producto total o productividad de la práctica lo suficiente, pero nada más que lo suficiente, para cubrir el coste de la mayor remuneración requerida para su traslado a la función j . En otras palabras, si P_j se redujera, aunque sólo fuese ligeramente, a fin de disponer de algún excedente para la redistribución, la función j dejaría de ser suficientemente atractiva para que se desplazasen hacia ella los individuos cuyas especiales cualidades o cuyo trabajo más duro dan como resultado el incremento del producto. En este caso, aunque Rawls no considerase que la práctica P era justa (porque la desigualdad sólo actúa en beneficio del hombre representativo j , y no en beneficio de todos), podríamos extender sus principios y decir que P es justa. Si hay un excedente de desigualdad, sin embargo —es decir, si el producto de la práctica con la remuneración incrementada a la posición j es más que suficiente para cubrir el incremento mínimo a la posición j , necesario para que a esa función se desplacen hombres cuyo trabajo dará como resultado un producto mayor—, entonces habrá muchas formas de distribuir ese excedente, ajenas a las valoraciones comparativas de Pareto. Cada una de ellas será "justa", según los principios de Rawls, en relación con la práctica sin la desigualdad en las remuneraciones; y cada una de ellas será *injusta* en relación con cada una de las otras. Si complicamos la situación aunque sólo sea ligeramente, considerando, por ejemplo, casos en los que el producto incrementado de la práctica requiera remuneraciones extra para dos o más funciones, las cosas se hacen, entonces, completamente inabordables. Y esto supone todavía que estamos comparando prácticas que son idénticas salvo en lo que se refiere a las remuneraciones.

Vale la pena de señalar aquí que la concepción inicial de Rawls de los dos principios sigue rigurosamente unas líneas *micro*-económicas. Tenemos que suponer, presumiblemente, que los cambios en las remuneraciones a las diversas funciones de la práctica no tendrán un efecto importante sobre la economía como conjunto. De ahí que podamos hablar fácilmente de redistribuciones de dinero dentro de la práctica, sin preocuparnos de la inflación ni de los cambios en los precios relativos. Si el principio de la diferencia hubiera de invocarse, no para una catalogación de prácticas en pequeña escala, sino para una evaluación de las grandes instituciones económicas básicas de toda la sociedad (como Rawls nos dice, en posteriores versiones del modelo, que ese modelo *debe* ser empleado), tendríamos que formular preguntas muy complejas, como la que nos permita indagar si una importante redistribución en el ingreso alterará los precios relativos hasta el punto de anular el supuesto efecto redistributivo, o, en otro caso, de hacer que las situaciones anterior y posterior a la redistribución no sean ya razonablemente comparables. Estos problemas llamados indicativos, que surgen cuando intentamos valorar el bienestar relativo de los individuos de dos etapas diferentes de la historia de una sociedad o de dos economías totalmente distintas, serán especialmente enojosos para Rawls, en cuanto se disponga a aplicar sus principios a ordenamientos institucionales macro-económicos.

¿SON LOS DOS PRINCIPIOS LA SOLUCIÓN DEL JUEGO DEL REGATEO?

Por último, debemos preguntarnos si los participantes en el juego del regateo de Rawls elegirían los dos principios que él ha propuesto, admitiendo que pudieran formarse un juicio coherente de ellos y acertar (de lo que nosotros hemos sido incapaces) con la forma en que habrían de utilizarse para catalogar unas prácticas alternativas, así como unas distribuciones alternativas a una práctica dada. La respuesta es simplemente "no", y las razones por las que esos principios no serían elegidos suponen mucho camino andando para llegar a explicar la invención de Rawls del velo de la ignorancia.

Empecemos centrando nuestra atención en el Principio IIb. Como he señalado, "abierto a todos" puede interpretarse, por lo menos, de varios modos plausibles diferentes, que incluyen "asignados sobre la base de una justa competición respecto a las habilidades o cualidades necesarias para realizar bien la tarea", y "asignados sobre la base de

un sorteo". En el primer caso, podemos suponer que los individuos más competentes tenderían a ocupar las posiciones que ofreciesen especiales remuneraciones (hablando en líneas generales, lo racional de la estratificación social en la moderna sociología funcionalista), mientras que, en el segundo caso, las posiciones se ocuparían al azar, yendo a parar las posiciones más codiciables a manos de los que hubieran ganado en el sorteo.

¿Qué ocurriría si se propusiesen para la adopción estas dos variantes del Principio IIb (suponiendo que los participantes en el juego se hubieran puesto de acuerdo, provisionalmente, sobre los Principios I y IIa)? A primera vista, podríamos suponer que todos optarían por la asignación mediante la competición. Es evidente que una sociedad regida por ese principio alcanzaría un máximo de eficacia, por lo que tendría la mayor riqueza total para distribuir, y mantendría así la posibilidad de prácticas unánimemente preferibles a cualesquiera otras alcanzadas mediante cualquier otro conjunto de normas fundamentales. Pero examinemos la cuestión más detenidamente (como harían, sin duda, los miembros de esa sociedad, pues son racionalmente egoístas).

Rawls ha estipulado que los individuos son "suficientemente iguales en poder y en habilidad para garantizar que, en circunstancias normales, ninguno puede dominar a los otros". Esto deja, sin embargo, un razonable grado de variación en el talento y en la ambición naturales, suficiente, en todo caso, para dar a unos individuos un margen de ventaja sobre otros, en una competición abierta por las posiciones favorecidas. Cada individuo tendría un conocimiento razonablemente claro de que sus talentos y capacidades se hallaban en función de los talentos y capacidades de los otros participantes. Unos comprenderían que sus facultades les situaban en posiciones más bien altas, y otros verían que se encontraban en posiciones relativamente desventajosas. Pero estos últimos se dirían: "Es cierto que la riqueza total de la sociedad será mayor si las posiciones se ocupan por competición; y si yo tuviese, en realidad, una probabilidad igual de alcanzar las tareas de la cumbre, encontraría razonable el acuerdo acerca de la versión competitiva del Principio IIb. Pero no es tan probable que yo alcance una de esas tareas, como que las alcancen mis conciudadanos de más talento; en efecto, como yo estoy entre los miembros de menos talento de esta sociedad, las probabilidades de que yo acabe en el fondo del montón en una competición justa son verdaderamente altas. En cambio, si los puestos se distribuyesen por sorteo, yo tendría mejores posibilidades de colocarme en esas tareas de la cumbre, aunque las remuneraciones

en ellas podrían ser menores, a causa de la reducida eficiencia de un sistema que coloca a gentes como yo en puesto de autoridad e importancia. Teniendo todo en cuenta (o, técnicamente, calculando el valor supuesto de las dos jugadas), mi racional egoísmo se inclina por la interpretación de la asignación por sorteo del Principio IIb¹³.

Los miembros más capaces de la sociedad efectuarían un cálculo análogo y llegarían a la conclusión directamente contraria de que su interés personal estaría mejor servido mediante la versión competitiva del Principio IIb. No hay un conjunto de normas con el que pueda comprometerse cada miembro de la sociedad, en la firme y racional esperanza de haber hecho lo mejor para sí mismo de cuanto razonablemente podía hacer.

Parecería, pues, que el regateo se habría hundido, simplemente, en un atolladero. Pero hay modos de salir de él. Como todos los individuos desean conseguir algún acuerdo, y como los dos principios, en una u otra versión de IIb, son unánimemente preferibles a no disponer de ningún medio, en absoluto, de resolver las disputas, sería irracional para uno y otro bando dejar que el regateo se rompiera. Ningún argumento racional puede persuadir a los menos capacitados de que deben, por su propio interés, adoptar la competición; y ningún argumento racional puede persuadir a los más capacitados de que deben, por su propio interés, aceptar la asignación por sorteo. La solución más apropiada, dentro de una teoría del juego, es, pues, la de recurrir al azar: una moneda al aire puede resolver la cuestión. Esto sería en beneficio de todos, porque si cada interpretación es unánimemente preferible a la falta de acuerdo, el promedio de las dos también lo será. Naturalmente, todo promedio ponderado de los dos, de la forma $[pIIb + (1 - p)IIb']$ también será unánimemente preferible a la falta de acuerdo, por lo que tal vez no debamos suponer demasiado fácilmente que el juego del regateo desembocará en un acuerdo unánime sobre un conjunto de principios.

Problemas análogos surgen en relación con la adopción del Principio IIa. Rawls afirma que los jugadores, en sus deliberaciones, estarán guiados por lo que se llama el principio maximin de decisión en la in-

¹³ Es de señalar, incidentalmente, que esta línea de razonamiento no viola, en modo alguno, el supuesto de la no-envidia. Los menos capacitados no envidian a los más capacitados la mayor remuneración que obtendrán en un sistema abierto al talento. Los menos capacitados, que llegan a tener una comprensión muy justa de los principios de la razón prudencial, hacen, sencillamente, un cálculo egoísta, no envidioso.

certidumbre¹⁴. "Estos principios expresan las condiciones de acuerdo con las cuales nadie se opone en absoluto éste es el principio maximin] a ver sus intereses limitados en el modelo de las prácticas, dados los intereses contrarios de los otros, en el supuesto de que los intereses de los otros se limiten también". Como Rawls señala, de un modo más bien pintoresco, las restricciones "pueden concebirse como las que una persona idearía, si estuviese diseñando una práctica en la que su propia situación habría de serle luego asignada por su enemigo".

Pero, ¿por qué los jugadores habían de elegir el principio maximin, extremadamente conservador, como guía en sus deliberaciones? La presunción de protegerse contra un enemigo, tomada literalmente, es inadecuada, toda vez que los jugadores no son envidiosos, y, por lo tanto, explicando ese supuesto en un sentido convenientemente general del término "envidia", no tienen interés alguno en rebajar a sus oponentes. Los menos capacitados podrían a primera vista, optar por el Principio IIa, sobre la base de que, como ellos eran los más probables candidatos a los puestos más bajos, no estaban dispuestos a aceptar una posibilidad de que los puestos inferiores estuviesen, realmente, peor pagados que en unas condiciones de absoluta igualdad de distribución. Los más capacitados, naturalmente, preferirían alguna forma de promedio del principio de utilidad, porque podrían esperar, razonablemente (sobre todo, si se había acordado la versión competitiva de IIb), que sus remuneraciones serían superiores al promedio, más bien que inferiores. En esta situación, con todos los jugadores haciendo los mejores cálculos posibles acerca de las consecuencias que razonablemente pueden esperar de la aplicación de los diferentes principios, pueden imaginarse toda clase de propuestas y de contrapropuestas, ninguna de las cuales parecería tener grandes posibilidades de conseguir la unanimidad en todas las partidas del juego del regateo. Por ejemplo, los menos aventajados podrían estar de acuerdo en una forma de utilidad media de IIa, la cual permitiría unas desigualdades que incrementarían la producción, al precio de un absoluto descenso del hombre representativo menos bien pagado, en lugar de una versión-sorteo del Principio IIb, que elevaría sus riesgos de acabar en las funciones menos beneficiadas.

En su "esbozo de una demostración", a mi parecer, Rawls cae en

¹⁴ Regla que nos dice que hay que maximizar la remuneración mínima de que disponemos. Ver más adelante, cuarta parte, y cualquier texto de la teoría del juego, como el de R. D. Luce y H. Raiffa, *Games and Decisions* (John Wiley & Sons, 1957).

un error lógico evidentemente simple. La cuestión no radica en que todos los participantes en el juego del regateo estén dispuestos a pasar de una práctica que ofrece un conjunto dado de remuneraciones a una nueva versión de la práctica que aumente las expectativas de todos los hombres representativos. La no-envidia y la positiva utilidad marginal del dinero lo garantizan. Podemos suponer, por lo tanto, que apoyarán el principio: "Elegir *siempre* una práctica (o una distribución en una práctica) unánimemente preferida, a una práctica o distribución dadas, cuando se ofrezca una". Pero de esto no se sigue, en absoluto, que apoyen el principio, mucho más discutible: "Elegir *sólo* una práctica o distribución en lugar de otra, cuando la primera sea unánimemente preferida a la segunda".

No es necesario escudriñar las insuficiencias de la primera forma del modelo de Rawls, sobre todo porque él mismo no tardó en abandonarla, por muchas de las razones que hemos venido examinando. Las dos principales fuentes de dificultades son, primera, la incapacidad de la relación de preferibilidad de Pareto para proporcionar un orden adecuado de prácticas alternativas, o de patrones alternativos de distribución dentro de una práctica; y, segunda, la imposibilidad de alcanzar la unanimidad entre un grupo de jugadores que, por así decirlo, saben demasiado acerca de sí mismos y de sus compañeros de juego. Técnicamente, el juego de Rawls es un *juego cooperativo de suma no-cero de n-personas*. Es de n personas, porque puede jugar cualquier número finito de personas; es de suma no-cero, porque las preferencias de los jugadores no son estrictamente competitivas (por ejemplo, todos los jugadores prefieren cualquiera de los posibles resultados antes que volver a la situación de partida o al resultado degradado de no lograr ningún acuerdo en absoluto); y es un juego cooperativo, en el sentido de que "los jugadores tienen una completa libertad de comunicación previa para establecer acuerdos *vinculantes* conjuntos"¹¹ (Luce y Raiffa, pág. 89; las cursivas son del original).

Rawls no nos dice cómo se juega realmente el juego, pero podemos suponer que las reglas son éstas: los jugadores se colocan arbitrariamente, de acuerdo con un orden determinado (por ejemplo, alfabéticamente); el primer jugador hace una jugada, que consiste en anunciar, con una voz que pueda ser oída claramente por todos, un principio o

¹¹ En ésta, como en otras cuestiones, me remito a R. D. Luce y H. Raiffa, *Games and Decisions*, y a T. Schelling, *The Strategy of Conflict* (Harvard University Press, 1960).

conjunto de principios para evaluar las prácticas a las que se comprometen o pueden comprometerse los jugadores, como miembros de la sociedad. El segundo jugador hace luego una jugada, que consiste en anunciar un principio o conjunto de principios. Los jugadores hacen jugadas de este tipo, siguiendo el orden de la lista de jugadores, una y otra vez, hasta que el árbitro anuncia que n jugadores consecutivos han hecho exactamente la misma propuesta (no comenzando necesariamente por el jugador número 1, desde luego). Entonces, el árbitro declara terminado el juego, y la propuesta que se ha formulado n veces seguidas se convierte en el conjunto de principios que, en adelante, los jugadores se comprometen a observar, en la evaluación de sus prácticas sociales y de sus instituciones.

Ahora, lo primero que es preciso señalar es que, como no hay límite de tiempo para el juego, y no se ha dicho nada, en absoluto, acerca de apremios temporales regidos por norma alguna (aburrimiento, la necesidad de recoger las cosechas sembradas, etc.), no hay razón de ningún tipo para que un jugador razonable ofrezca un principio de compromiso. Como las jugadas son gratuitas, lo único razonable que cada jugador puede hacer es proponer: "Que, de hoy en adelante, todo se haga exactamente como yo deseo". Hay una probabilidad no-cero de que todos los otros $n - 1$ jugadores acepten, estúpidamente, y, en ese caso, el que haya hecho la propuesta habrá ganado el juego, evidentemente.

Pero si se impone un límite de tiempo se destruye la simetría del juego, porque la última persona que realiza una jugada antes de que el tiempo acabe se halla en una situación distinta de aquella en que se encuentra la penúltima, y así sucesivamente. Como Thomas Schelling ha señalado, una vez que abandonamos el supuesto de la simetría, resulta vitalmente importante determinar el lugar exacto que usted ocupa en el orden de los jugadores, en relación con el momento límite del tiempo. En el juego de Rawls supongamos que cada jugada (cada anuncio de un conjunto propuesto de principios) consume un determinado tiempo mínimo finito = t segundos. El jugador que realiza su jugada exactamente nt segundos antes del tiempo límite cuenta con una enorme ventaja sobre todos los demás. Puede proponer un conjunto de principios que, si bien los otros pueden temer razonablemente que le beneficien de un modo indebido, son de tal género, sin embargo, que a todos los jugadores les interesará aceptarlos, antes que retroceder a la situación de "estado de naturaleza" de no ponerse de acuerdo sobre ninguna regla en absoluto. Los jugadores restantes se encuentran ahora en

desventaja. Si alguno de ellos no acierta, en esta última ronda, a enunciar los principios del jugador favorecido, ya no habrá tiempo suficiente para establecer ningún otro conjunto de principios, con lo que caerán nuevamente en un estado de naturaleza. Pero como todos son razonables y no envidiosos, aceptarán la propuesta formulada, de igual modo que lamentarán el hecho de que el jugador favorecido tenga sobre ellos una ventaja de negociación tan considerable. (Ver Schelling, Apéndice B, especialmente págs. 274-75, nota 11.)

Las dificultades que acabamos de esbozar podrían superarse cambiando ese juego por un juego no cooperativo, en el que "no se permita absolutamente ninguna comunicación previa entre los jugadores" (Luce y Raiffa, pág. 89). El juego se convertiría, entonces, en lo que se llama un juego de coordinación: cada jugador escribiría su propuesta de principios en una hoja de papel y la entregaría al árbitro; si todas las hojas de papel contuviesen exactamente los mismos principios, éstos se convertirían en las normas de la sociedad; en otro caso, la sociedad retrocedería al estado de naturaleza. Schelling ha explorado esos juegos de coordinación con gran imaginación e inteligencia, y demuestra muy convincentemente (en mi opinión) que el éxito de la coordinación depende de la influencia de factores ajenos al modelo puramente formal del juego, y, en el sentido teórico del juego, irracionales (como, por ejemplo, la existencia de tradiciones culturales que todos los participantes conocen y saben que los otros participantes conocen, o incluso la presencia de un letrado destacadamente expuesto en el lugar de la negociación con el lema: "compartir y compartir por igual"). En la primera forma del juego, en la que los jugadores saben quiénes son y qué posiciones ocupan en la sociedad, no hay base alguna, en absoluto, para suponer que un juego de coordinación desembocará en un conjunto de principios unánimemente elegido, y menos razón todavía, si ello fuera posible, para suponer que el resultado de cada partida del juego será un conjunto de principios y sólo uno. En la forma final del modelo, una vez que se ha tendido el "velo de la ignorancia" (sobre esto volveremos luego), la situación elegida ha cambiado tanto, que Rawls tiene que recurrir a consideraciones totalmente diferentes para defender su pretensión de que sus dos principios de la justicia constituyen la solución del juego.

Por último, aquí puede ser conveniente una palabra acerca de la regla maximin de elección, aun cuando el tema se tratará detenidamente en la cuarta parte. En los juegos de suma-cero de dos personas, resulta que hay argumentos muy brillantes (aunque de ningún modo incon-

trovertibles) que aconsejan calcular lo peor que puede ocurrirle a usted como consecuencia de cada estrategia o línea de juego disponible, y elegir después la estrategia que ofrece el "mejor peor". El hacer máxima su remuneración mínima, o maximin, tiene una variedad de propiedades matemáticas en los juegos de suma-cero de dos personas que lo recomiendan como un procedimiento racional. Como los juegos de suma-cero de dos personas tienen, desde luego, la característica de que todo lo que usted gana su adversario lo pierde, y viceversa, hay una cierta conveniencia literal en jugar como si su peor enemigo fuese a determinar, dentro de los límites del poder que él tiene, lo que le ha ocurrido a usted. Pero en los juegos de suma no-cero de n -personas, no puede darse significado alguno a la noción de que un jugador gana lo que el otro jugador pierde. (Nota: a causa de la forma en que se definen las funciones de utilidad, no es posible, por ejemplo, distinguir entre juegos de suma no-cero en los que "todos pueden ganar" y juegos de suma no-cero en los que "todos pueden perder". El rasgo conocido como "condición de suma-cero" no es más que una consecuencia del carácter estrictamente competitivo de las preferencias de los jugadores y de la permisibilidad de transformaciones lineales de las funciones de utilidad).

Por lo tanto, no es obvio, en absoluto, ni matemática ni filosóficamente, que la regla maximin sea la regla de elección racional. Hay argumentos perfectamente válidos para esa elección a fin de maximizar la utilidad promedio, y si algún jugador considera esto como una fruslería arriesgada, siempre puede valorar sus promedios para expresar su grado de aversión al riesgo (además del valor que ha dado ya a los diversos paquetes de remuneraciones en la construcción de su función de utilidad). Ni en la versión cooperativa ni en la no-cooperativa del juego de Rawls hay razón alguna para suponer que unos jugadores racionales hayan de alcanzar siempre un acuerdo o una coordinación respecto a los dos "principios de la justicia".

VI

LA SEGUNDA FORMA DEL MODELO

LLEGADO A este punto, Rawls introduce dos importantes alteraciones en su teoría, con el expreso propósito, evidentemente, de responder a los tipos de objeciones que hemos indagado en nuestro examen de la primera forma del modelo. Los dos cambios son, primero, el problemático recurso del velo de la ignorancia, y, segundo, la sustitución del principio simple de la preferibilidad de Pareto por el principio de diferencia, en su forma de "hombre representativo menos aventajado".

Antes de examinar estas dos innovaciones teóricas, acaso sea conveniente que nos detengamos, por un momento, a señalar una característica un tanto extraña de la exposición de Rawls. Muy pronto, ya en 1958 o antes, Rawls había optado por una fórmula para expresar sus "dos principios de la justicia". A pesar de ciertas confusiones, que hemos examinado ya, los principios, tal como se exponían, tenían una interpretación perfectamente natural para cualquier lector corriente de inglés. En términos generales, aseguraban que una sociedad debía elegir como línea de partida el punto de distribución igual más próximo a su límite de producción, y apartarse sólo de ese punto para trasladarse a algún otro punto —Pareto prefería el de partida— en el que las posiciones relativamente favorecidas estuviesen abiertas a todos, en equitativa competencia. Rawls pretendía que estos principios eran, simultáneamente, una fiel reconstrucción de nuestras convicciones morales establecidas acerca de materias de justicia institucional (después volveremos sobre esto), y la solución a un juego del regateo definido de tal modo que capte la noción esencial de una sociedad de individuos racionalmente interesados, dispuestos a comprometerse con un conjunto de principios.

Cuando las ambigüedades, insuficiencias y contradicciones de estas dos aseveraciones acerca de sus principios resultan evidentes para Rawls, su reacción inmediata es la de abandonar su fórmula, y buscar, en lugar de ella, un conjunto diferente de principios que satisfagan las exigencias teóricas que él quiere plantearles. En resumen, si la primera

solución propuesta a un juego del regateo resulta errónea, hay que buscar otra. Pero Rawls adopta una actitud totalmente distinta. Trata la formulación original de los dos principios como dada, como *ex cathedra* por así decirlo, y luego se lanza a la búsqueda de alguna forma de "interpretación" de las palabras, de modo que ofrezcan un sentido defendible. Así, en "Distributive Justice" (Justicia distributiva)¹⁶, aun cuando ha alterado sustancialmente su teoría, restablece los principios en un lenguaje que, excepto en la sustitución de "prácticas" por "instituciones" y en unas pocas enmiendas gramaticales, es literalmente idéntico al de la formulación de "Justice as Fairness" (La justicia como imparcialidad). Ahora dice, como si estuviera tratando de descifrar el texto de un rompecabezas, que "no está claro lo que se quiere significar al decir que las desigualdades deben actuar en favor de todos los hombres representativos". Primero, sugiere que se compare la institución sometida a estudio con alguna "referencia históricamente importante". Luego, propone directamente la optimalidad de Pareto como el "significado". Por último, encuentra "una tercera interpretación", concretamente la maximización de las expectativas del hombre representativo menos aventajado. "Interpretaciones" similares conducen después a Rawls al principio de la prioridad de la libertad y a otros cambios teóricos del modelo original.

Ahora bien: es perfectamente razonable que tratemos de este modo un conjunto de palabras, si nos llegan con una cierta impronta de autoridad — por ejemplo, si son las palabras de Dios. Cuando la razón entra en conflicto con el evidente significado de los mandamientos de Dios, una respuesta natural (aunque no la única) consiste en interpretar el mandamiento. No podemos cambiarlo simplemente, porque es la palabra de Dios, pero podemos reflexionar en busca de un significado que esté conforme con la razón y con la intuición moral. El mismo tipo de procedimiento tiene un cierto sentido cuando se trata de la obra de un gran pensador, que, sin embargo, resulte oscuro. Algunos filósofos, a veces, tienen profundas concepciones, ricas en sugerencias y en expectativas, pero cuya expresión es oscura o confusa. Aventuramos nuestro tiempo y nuestra energía, por así decirlo, cuando buscamos una interpretación de sus palabras que sea coherente con su orientación teórica básica, y que tenga, sin embargo, un sentido filosófico sostenible. Pero resulta ligeramente extraño que un pensador trate de este

modo sus propias palabras. Queremos decir que Rawls, seguramente, debería decidir que su solución inicial al juego del regateo era correcta y defenderla, o, en otro caso, decidir que era errónea y cambiarla. Pero, ¿por qué decidir que era errónea, y luego optar por la vía intermedia de mantener las palabras y cambiar su significado?

Al fin, aunque con lo que yo sólo puedo considerar como una gran renuencia, Rawls revisa la auténtica exposición verbal de sus dos principios. Pero este estilo profundamente conservador de desarrollo intelectual es revelador e importante como orientación para la estructura del pensamiento de Rawls. Es sumamente instructivo comparar a Rawls con otro pensador moral y social de la tradición angloamericana, John Stuart Mill. Durante toda su larga vida, Mill luchó por liberarse de los estrechos límites conceptuales de la simple fe utilitarista en que le educaron su padre, James Mill, y su padrino, Jeremy Bentham. En libros como sus *Principios de Economía política*, podemos ver a Mill considerando, teniendo en cuenta, reconociendo los méritos de las escuelas de pensamiento más variadas y competitivas, desde el romanticismo y el idealismo hasta el socialismo utópico. Casi no puede formularse crítica alguna contra Mill, que él no haya discutido e incluso aceptado. Pero, a pesar de todo, sigue siendo, fundamentalmente, un utilitarista.

A mi parecer, Rawls muestra un sistema de pensamientos muy similar a éste. En el comienzo del desarrollo de su filosofía, encontró el recurso del juego del regateo y los dos principios como solución. Se mantuvo adherido a ellos, a través de todas las revisiones, complicaciones y acomodaciones de la teoría, y ese sistema es el que domina la exposición de *Teoría de la justicia*, a pesar del tono extremadamente modesto y condescendiente del último capítulo. Más adelante sugeriré un buen número de razones para dudar de que los dos principios, en cualquiera de sus formas, sean elegidos por individuos que se encuentren en cualquiera de las diversas posiciones originales bosquejadas por Rawls. Pero me parece justo decir que Rawls ha pensado, y pensado detenidamente, en la mayor parte de las objeciones que yo y otros críticos formulamos, aunque a veces no responda a las objeciones ni altere su teoría para ajustarse a ellas.

¹⁶ Laslett y Runciman, *Philosophy, Politics, and Society* (Barnes & Noble, Inc., 1967); tercera serie, págs. 58-82.

EL VELO DE LA IGNORANCIA

El velo de la ignorancia aparece por primera vez en "Distributive Justice". Al describir la situación de los jugadores en el juego del regateo, Rawls dice que es preciso imaginarlos colocados

"en una situación inicial debidamente definida, uno de cuyos rasgos importantes consiste en que ninguno conoce su posición en la sociedad, ni siquiera su lugar en la distribución de talentos y facultades naturales. Los principios de la justicia a los que todos están ligados para siempre se eligen en ausencia de este tipo de información específica. Un velo de ignorancia impide que cualquiera se vea favorecido o perjudicado por las contingencias de clase social y de fortuna; y de ahí que los problemas de regateo que surgen en la vida cotidiana a causa de la posesión de este conocimiento no influyan en la elección de los principios."

No es necesario decir que el velo de la ignorancia constituye un recurso analítico, y no una propuesta utópica ni una nota de una ingenuidad elemental. La consecuencia del recurso es la de establecer que, en sus razonamientos, los participantes en el juego del regateo tienen que abstraerse de las particularidades de su fortuna o de sus facultades naturales y sociales, y, por consiguiente, no deben tenerlas en cuenta. La restricción es análoga a la exigencia, en geometría pura, de que se haga abstracción de las características particulares de la figura que se estudia, cuando se trata de demostrar un teorema, y se atiende sólo a algunas de sus propiedades matemáticas. En realidad, como Rawls piensa que el razonamiento de los jugadores no es deductivo, sino, en algún sentido, constructivo, podemos decir que impone a sus jugadores el mismo tipo de restricción que Kant pensaba que nos era impuesto por la forma de intuición pura en nuestro razonamiento matemático puro.

El velo de la ignorancia tiene muchos alicientes, desde el punto de vista de Rawls. El primero y más evidente es el de que desbarata algunas de las más poderosas objeciones a la primera forma del modelo. Como hemos visto, las dos razones principales por las que no podía esperarse que unos jugadores racionales eligiesen los dos principios de Rawls, eran, primero, la de que su conocimiento de sus propios talentos y facultades especiales les conduciría a discrepar acerca de un principio que asignase a los individuos unas posiciones desigualmente remuneradas, y, segundo, la de que este mismo conocimiento conduciría a los más capacitados a defender unos principios que permitiesen ligeras reducciones en los ingresos de los menos favorecidos, a cambio de

sustanciales incrementos en todas las otras posiciones. He sugerido que la imposibilidad de alcanzar un acuerdo no desembocaría en un acuerdo unánime sobre los dos principios de Rawls, como único medio de evitar una ruptura del regateo, sino más bien en alguna de las numerosas combinaciones de principios —mixtas o casualmente seleccionadas—, que favorecen a diferentes sectores de la sociedad de jugadores. Pero el velo de la ignorancia niega a los jugadores, precisamente, la información que les conduciría a ese tipo de desacuerdos. Por eso ofrece alguna esperanza de transformar el juego del regateo en un juego que tenga una *solución*, en el sentido técnico preciso.

El velo de la ignorancia tiene también alicientes positivos, y éstos, evidentemente, son muy importantes para Rawls. En primer lugar, al negar a los jugadores toda información acerca de su "posición en la distribución de talentos y facultades naturales", les inclina hacia la adopción de un principio que trate esos talentos y esas facultades como recursos sociales, más que como recursos personales. Al no saber en qué lugar se encuentra en la jerarquía de la distribución, cada jugador —en opinión de Rawls— buscará un principio que coloque el fruto de esos talentos y facultades a disposición de los miembros de la sociedad en general. El resultado neto será el de eliminar lo que Rawls y muchos otros pensadores morales consideran la injusticia social de remunerar a los individuos por el azar de su posesión de talentos naturales económicamente provechosos.

El velo de la ignorancia tiene también el efecto de obligar a los jugadores a adoptar un punto de vista generalizado, que guarda una gran semejanza con lo que algunos pensadores morales llaman "el punto de vista moral"¹⁷. Los hechos que el velo niega a los jugadores son precisamente aquellos hechos que parece inadecuado tener en cuenta cuando se hacen juicios morales, como opuestos a juicios puramente prudenciales. Si Rawls puede demostrar que un grupo de agentes prudencialmente racionales, cada uno de los cuales busca su interés racional propio, llegaría, ignorando las particularidades de sus condiciones individuales, precisamente al conjunto de principios que sería apoyado desde un punto de vista moral, habrá realizado un importante avance teórico. En realidad, habrá demostrado que pueden alcanzarse conclusiones morales, sin abandonar el punto de partida prudencial y proponiendo un enfoque moral, simplemente prosiguiendo el razonamiento

¹⁷ En la tercera parte de este ensayo, discutiré detenidamente la llamada interpretación kantiana de la posición original y su relación con el velo de la ignorancia.

prudencial propio, bajo ciertas restricciones procedimentales de regateo y de conocimiento. Así, la maniobra promete delimitar el interminable e indeciso argüir y contra-argüir de las distintas escuelas competitivas de la teoría moral. Como Rawls debe de haber apreciado la fuerza de la idea, y como, en principio, ésta parecía resolver los problemas de la primera forma del modelo, no es difícil comprender por qué adoptó el velo de la ignorancia como una modificación de su teoría.

EL PRINCIPIO DE DIFERENCIA

El segundo cambio importante fue la reelaboración del Principio IIa en la forma de lo que ha pasado a llamarse "el principio de diferencia". Rawls lo introduce, un tanto disimuladamente, como la tercera "interpretación" posible del lenguaje de los dos principios originales:

"Hay, sin embargo, una tercera interpretación que es inmediatamente sugerida por las observaciones anteriores [relativas a las insuficiencias de la optimalidad de Pareto como principio], y ésta consiste en elegir alguna posición social con referencia a la cual ha de ser juzgado el patrón de expectativas como conjunto, y luego maximizar con respecto a las expectativas de este hombre representativo, sin perjuicio de las demandas de libertad igual y de igualdad de oportunidades. Ahora bien: el único candidato evidente es el hombre representativo de los menos favorecidos por el sistema de desigualdades institucionales".

El recurso a la versión "menos aventajada" del principio de diferencia resuelve inmediatamente dos de los más graves problemas planteados por la primera forma del modelo. Ante todo, permite un completo ordenamiento, y no sólo un cuasi-ordenamiento, de series alternativas de prácticas, instituciones, conjuntos de instituciones, sociedades, etc. Como entre dos instituciones, se identifica sencillamente al hombre representativo menos aventajado bajo cada una, y se comparan las remuneraciones que se "le" asignan (sin que sean, desde luego, necesariamente el mismo individuo o grupo de individuos en cada caso). Debe favorecerse la institución que asigne una mayor remuneración al hombre menos aventajado, independientemente de cuál pueda ser el resultado total, o el bienestar total, o el patrón de otras remuneraciones. Como en dos instituciones cualesquiera, el hombre representativo menos aventajado de la primera estará mejor, igual, o peor que el hombre menos aventajado de la segunda, y como este sentido de

"mejor que" es transitivo, la nueva versión del principio de diferencia definirá un completo y débil ordenamiento¹⁸.

La segunda virtud del nuevo principio de diferencia consiste en que disuelve el rompecabezas de comparar prácticas con diferentes números de funciones, definidas de modo que no permiten una comparación natural o fácil. Cada práctica tendrá que definir una función menos aventajada, aunque sea una práctica con una sola función. Desde luego, este cambio nos permitirá establecer algunas comparaciones más bien extrañas, como, por ejemplo, una comparación entre el matrimonio monógamo y el sistema de instalación Volvo de fabricación de automóviles en serie. Teóricamente, no hay nada en el principio de Rawls que nos impida preguntar qué práctica es preferible y responder luego comparando al hombre representativo menos aventajado del sistema Volvo con el hombre representativo menos aventajado (probablemente, la mujer) en la práctica del matrimonio monógamo. Si nos sentimos incómodos al tener que elegir entre el coche familiar y la familia, siempre podemos definir alguna tercera práctica —resultado de la combinación de las dos— en la que el hombre representativo menos aventajado esté mejor que en cualquiera de las otras dos, separadamente.

¹⁸ Una vez más, es necesario matizar que no se están teniendo en cuenta las variaciones en los precios o cambios relativos en la identidad y naturaleza de las mercancías. Como estamos comparando las remuneraciones de los "hombres representativos" en diferentes prácticas, y como la identidad del hombre representativo menos aventajado puede cambiar de una práctica a otra, Rawls tiene que desplazarse de consideraciones de utilidad a consideraciones de remuneraciones intersubjetivamente comparables, ya sean de dinero o de paquetes de mercancías.

VII

CRITICA DE LA SEGUNDA FORMA DEL MODELO

EL VELO de la ignorancia y la nueva forma del principio de diferencia marcan un importante avance sobre la primera forma del modelo, pero, tras un conveniente examen, inmediatamente comprobamos, como Rawls debe de haber hecho, que son necesarios nuevos cambios y elaboraciones antes de que este nuevo modelo pueda servir a los objetivos que Rawls se propone.

La primera dificultad, naturalmente, consiste en que, a menos que especifiquemos con exactitud todo lo que el velo de la ignorancia oscurece y todo lo que deja disponible para los ocupantes de la posición original, no podremos deducir nada, en absoluto, acerca de los procesos de razonamiento de éstos. Algo similar habrá que decir de las clases de general conocimiento o de particular información que los jugadores retienen, una vez que han entrado en el juego del regateo y han olvidado quiénes son en particular.

Un segundo problema, potencialmente más grave que este primero, es el planteado por el velo de la ignorancia. En el modelo original, se supone que los jugadores son individuos racionalmente egoístas, con aspiraciones, intereses, objetivos y planes totalmente desarrollados y conscientemente definidos. Pero, siguiendo la huella del utilitarismo clásico y de la moderna economía del bienestar, Rawls no estableció supuestos sustanciales acerca del carácter de aquellos objetos o intereses. Por ejemplo, no se suponía que valoraban los deleites del espíritu por encima de los placeres del cuerpo, o que estimaban o despreciaban la compañía de su prójimo. La única coacción de la que se suponía que actuaba sobre sus preferencias era el supuesto común de una utilidad marginal positiva de cualesquiera bienes que la práctica en cuestión estuviese distribuyendo¹⁹.

¹⁹ Con la ulterior condición anti-trivializante de que el término "bienes" tenía que ser comprendido objetiva y socialmente, más que subjetiva e individualmente. Para el punto de vista contrario, ver I. M. D. Little, *A Critique of Welfare Economics* 2.^a ed., Oxford at Clarendon Press, 1957, pág. 16.

Bajo el velo de la ignorancia, sin embargo, los jugadores no tienen idea alguna de sus propósitos, planes e intereses. Parecería, pues, que no pueden decir nada, en absoluto, acerca de la forma en que les gustaría que se ordenasen sus instituciones. Naturalmente, cada uno de ellos puede manifestar alguna aspiración general y vaga, como la de que, cuando descubren qué es lo que necesitan, desean que la práctica de la que ellos forman parte asigne a su función en ella algún conjunto de remuneraciones que les ayude a conseguir lo que necesitan. Pero en el regateo que ha de determinar la elección de los principios constructivos para la evaluación de las prácticas sociales, es difícil ver cómo los jugadores podrán razonar mientras sean completamente ignorantes de toda realidad sustancial de sus deseos y propósitos. Algo así habrá que decir también de las clases de propósitos que los jugadores tienen en sus existencias reales, aunque, naturalmente, nada pueda decirse que diferencie a los unos de los otros o que les permita inferir, de alguna forma, sus verdaderos talentos, facultades o intereses.

El principio de diferencia también plantea algunos problemas nuevos, incluso mientras resuelve los viejos. El primer problema se refiere al alcance del principio, al tipo de los casos a cuya aplicación está destinado. En el modelo original de su teoría, Rawls hablaba mucho acerca de funciones y de juegos, de un modo que sugería una idea más bien de pequeña escala de la justicia de las prácticas. Nos sentíamos impulsados a pensar en el juego del béisbol, o en la organización de una fábrica, o en la estructura de autoridad jerárquica de un ejército. Como he observado varias veces, Rawls parecía suponer implícitamente que su principio sería aplicado a unas prácticas, cada una de las cuales era de una escala suficientemente pequeña, en relación con la sociedad de que formaba parte, que los cambios en su patrón de remuneraciones tendrían unos efectos insignificantes sobre la economía total de la sociedad. Mientras se podía imaginar a los jugadores como individuos identificables que regateaban en plena posesión de los detalles de su situación, esta concepción del modelo de Rawls tenía un sentido. Pero, una vez que pasamos a la versión del velo de la ignorancia, en la que se imagina que los jugadores adoptan una perspectiva impersonal, y tal vez incluso intemporal, la interpretación de pequeña escala o micro-interpretación de los dos principios parece inadecuada. Es más aceptable construir el juego del regateo como determinante, en todo momento, de la estructura más fundamental de la economía y de la sociedad. En la forma inmediata del modelo, consecuentemente, Rawls limita, de un modo explícito, el principio de diferencia a ordenamientos

macro-económicos y macro-sociales. Esto, a su vez, genera, como veremos, algunos problemas extraordinariamente complejos.

Una segunda dificultad, que ha sido muy señalada y comentada por recientes críticos de Rawls, se refiere a algunas de las implicaciones contra-intuitivas del principio de diferencia revisado. En la forma de preferencia de Pareto original, el principio definía un cuasi-ordenamiento que, mientras se mantuvo, fue muy atractivo como norma de elección colectiva. Como se exigía que las desigualdades operasen en beneficio de cada uno, se podía estar seguro de que una práctica considerada justa por el Principio IIa obtendría una aprobación unánime. El nuevo principio, desde luego, establece un ordenamiento completo, más que un simple cuasi-ordenamiento, pero, al hacerlo así, compromete a los jugadores con juicios seriamente cuestionables de preferibilidad relativa. Como esta objeción ha sido formulada por muchos críticos de Rawls y (significativamente) es reconocida por el propio Rawls, será suficiente un solo ejemplo para revelar la naturaleza del problema²⁰.

Imaginemos un sistema social de instituciones y prácticas económicas, en el que un cierto número de tareas productivas funcionalmente diferenciadas están diferencialmente remuneradas, pero en el que la distancia entre los trabajos mejor y peor remunerados no es muy grande. Imaginemos (si realmente fuera así) que la reducida distancia entre los diferentes niveles de remuneración tiene como consecuencia que no haya una suficiente aparición de talentos. Como la remuneración de los trabajos más altamente calificados no es mucho mejor que la de los trabajos sin calificar, los individuos no se sienten fuertemente impulsados a someterse a la preparación adicional necesaria para adquirir las calificaciones requeridas. Más aun: algunos tipos de trabajos muy altamente especializados ni siquiera existen en el sistema, porque nadie está dispuesto a llevar a cabo el enorme esfuerzo personal necesario a fin de prepararse para ellos. Supongamos ahora que se propone una revisión de este sistema de producción y distribución, que implica una estructura laboral más altamente diferenciada, con una amplia comunidad ascendente de calificaciones laborales, y que requiere para su eficaz funcionamiento una amplitud mucho mayor en la estructura de las

²⁰ Ver, por ejemplo, las primeras secciones del ensayo de John C. Harsanyi, "Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls' Theory" *APSR*, Vol. 69, núm. 2 (junio, 1975), págs. 304-606. Ver también la discusión, muy elaborada, de Douglas Rae, "Maximin Justice and an Alternative Principle of General Advantage", *ibid.*, págs. 630-47.

remuneraciones. En este sistema revisado, todo hombre representativo, *excepto el menos acomodado*, está considerablemente mejor que antes (suponiendo que pueda hacerse una aceptable comparación entre los sistemas, dado el cambio en el número de funciones), pero el hombre representativo menos acomodado está ligeramente peor. Por último —y esto es muy importante—, supongamos que *no* hay ningún excedente de desigualdad del que puedan pagarse transferencias a la clase menos aventajada. El sistema requiere una igualación tan sensible de talentos y trabajos, que si de los trabajos mejores se tomase algo importante en el concepto de las remuneraciones para compensar a los menos aventajados, la eficacia del sistema se reduciría hasta un punto en el que la comunidad estaría peor que antes. Según Rawls, las personas racionales en la posición original adoptarían un principio (el principio de diferencia) que excluiría el desplazamiento hacia este sistema revisado desde el sistema original, independientemente de la magnitud de las ganancias para todas las clases excepto para la menos aventajada, independientemente del carácter marginal de la desventaja para los menos aventajados, e independientemente de la exigüidad del volumen de esa clase menos aventajada²¹. Rawls necesitará alguna norma muy poderosa y aceptable de elección racional para justificar una selección de principios a lo largo de esas líneas en la posición original. Su propuesta, como veremos en seguida, es la llamada regla maximin de elección bajo incertidumbre.

Los cuatro problemas que acabamos de examinar, dos de los cuales surgen del velo de la ignorancia y los otros dos de la nueva versión del principio de diferencia, son tratados por Rawls a través de una serie de importantes revisiones, adiciones y alteraciones, cuyo resultado es una forma distintivamente nueva del modelo.

²¹ No puede Rawls protestar que los cálculos del tipo que yo he imaginado exceden del poder del más elaborado modelo econométrico. Eso es cierto, pero lo mismo puede decirse de los tipos de cálculos exigidos por su principio.

VIII

LA TERCERA FORMA DEL MODELO

COMO HEMOS visto, el velo de la ignorancia y la nueva versión del principio de diferencia planteaban a Rawls cuatro problemas: primero, no estaba claro qué era, suponiendo que fuese algo, lo que había que pensar que los jugadores en la posición original conocían acerca de su propio mundo; segundo, no estaba claro cómo podían entregarse a una deliberación racional, en ausencia de toda concepción de sus objetivos e intereses; tercero, el alcance de la aplicación de los principios sobre los que estaban decidiendo se encontraba bastante mal definido; y, por último, las implicaciones claramente contra-intuitivas del nuevo principio de diferencia exigían alguna descripción un tanto más adecuada de los principios de elección racional a los que podía esperarse que ellos recurriesen. En respuesta a estos cuatro problemas, Rawls formula una serie de elaboraciones teóricas que, tomadas en su conjunto, nos acercan mucho a una versión a gran escala de la teoría, tal como aparece en *Teoría de la justicia*.

Las "partes" en la posición original saben mucho, aunque no acerca de sí mismas en particular. (En este punto, Rawls deja de referirse a las personas en la posición original como a "jugadores" en un juego de regateo. El juego, por así decirlo, ha sido llamado a cuenta de la ignorancia.) Lo único en particular que ellos saben es que "su sociedad está sometida a las circunstancias de la justicia", es decir, a una moderada escasez, a una aproximada igualdad de poder, etc. No está claro por qué es necesario que sepan que su sociedad es de tal carácter que en ella se planteará la cuestión de la justicia. Acaso bastaría preguntarles qué principios elegirían, si su sociedad fuese de ese carácter. Pero el tema no es importante, a no ser como expresión del grado en que Rawls continúa manteniendo viva la imagen de un juego del regateo, aun después de que ha negado los supuestos previos que dan sentido lógico a ese concepto.

Pero lo que *es* importante es que las partes "conocen los hechos ge-

nerales acerca de la sociedad humana" (164)²². Según Rawls, "entienden las cuestiones políticas y los principios de la teoría económica; conocen las bases de la organización social y las leyes de la psicología humana" (164).

En la cuarta parte, indicaré algunas razones para creer que la hipótesis de Rawls es imposible. Las partes no podían poseer, en principio, el conocimiento total y único que Rawls les atribuye. Más aun: la imposibilidad es lógica, no simplemente genética. Pero está bastante claro a qué aspira Rawls. Hay que imaginar a las partes en la posición original como análogas a entidades de un sistema físico que saben que poseen masa, que se encuentran en una posición, y que están (o no están) moviéndose respecto al sistema, con una determinada velocidad en una determinada dirección; y que conocen las leyes del movimiento que rigen el sistema, incluidas ellas mismas; pero que no saben lo que realmente son su masa, su posición y su velocidad, y, por lo tanto, no pueden inferir nada en particular acerca de sus características físicas. Sabiendo todo esto, podrían inferir cualquier número de proposiciones hipotéticas, de la forma: "Si yo tengo una masa m , en una posición p , y me muevo respecto al sistema con una velocidad v , y si choco con otra masa m' , etc., entonces, después del choque, me trasladaré con una velocidad v' , etc."

Desgraciadamente para Rawls, aun cuando fuese posible, en principio, un conocimiento no específico de este tipo acerca de la sociedad, sería muy difícil, en este momento del desarrollo de las ciencias sociales, decir en qué consistiría. ¿Cuáles son los principios de la teoría económica? ¿La ortodoxia neoclásica? ¿Antes o después de la controversia principal? ¿Con o sin funciones de seudoproducción? ¿Es un "principio de teoría económica", o sólo un hecho casual, que la razón capital/producto, en las modernas economías industriales, varíe muy poco del valor 3:1? ¿Qué se entiende por "las bases de la organización social"? ¿Sólo unas pocas y muy generales verdades acerca de la necesidad de diferenciación e integración funcionales, o algo más estrechamente relacionado con ella, como la ley de hierro de la oligarquía, de Michel? Respecto a las "leyes de la psicología humana", a las que Rawls apela una y otra vez en las fases ulteriores del desarrollo de su teoría completa, la categoría de las mismas es realmente muy incier-

²² Las referencias entre paréntesis remiten a *A Theory of Justice*. (En esta versión castellana, remiten a *Teoría de la justicia*, Fondo de Cultura Económica, 1979. Me abstengo, pues, de traducir esas referencias, limitándome a transcribirlas literalmente [N. del T.])

ta. Supongo que Piaget se sentiría feliz si viese sus teorías de desarrollo cognoscitivo elevadas a la categoría de leyes; Erik Erikson podría sentir una salvadora turbación ante los intentos de tratar su concepción del ciclo vital, de un modo semejante; pero Freud se habría opuesto, seguramente, a toda sugerencia de que la incesante búsqueda del subconsciente pudiera transformarse en un conjunto de leyes teóricas cuyo significado podría conocerse con abstracción de las particularidades de la vida del conocedor.

En todo caso, respecto a las partes en la posición original, "se da por sentado que conocen todos los hechos generales que atañen a la elección de los principios de justicia". Pero aún no están preparados para comenzar sus deliberaciones, porque todavía no saben nada acerca de sus razones para deliberar. A fin de aplicar los principios de una razón prudencial al problema de elección con que se enfrentan, deben dar por sentado algo, por poco que sea, acerca de lo que se proponen al buscar una solución. A fin de descarnar la situación de elección de las partes lo suficiente para ponerlas en marcha, desde el punto de vista deliberativo, Rawls establece otros dos supuestos, cada uno de ellos muy poderoso y altamente discutible.

Para empezar, Rawls atribuye a cada parte el conocimiento general de que los seres humanos racionales, en la medida en que son racionales, formulan o, por lo menos, actúan como si hubiesen formulado lo que puede llamarse "planes de vida". Introduce el concepto al comienzo de su exposición de los principios de justicia, y vuelve a él mucho después, como parte de la discusión más general de "La bondad como racionalidad". El pasaje siguiente capta muy bien las ideas:

"La idea principal es que el bien de una persona está determinado por lo que para ella es el plan de vida más racional a largo plazo, en circunstancias razonables. Un hombre es feliz en la medida en que tiene mayor o menor éxito en llevar a cabo este plan. Para decirlo brevemente: el bien es la satisfacción del deseo racional. Hemos de suponer entonces que cada individuo tiene un plan racional de vida, diseñado según las condiciones a que se enfrentan, y también para permitir la satisfacción armónica de sus intereses. Sus actividades están programadas de modo que se puedan satisfacer sin interferencias los diversos deseos. Se llega a él rechazando otros planes que, o bien es menos probable que tengan éxito, o bien no hacen posible una realización tan amplia de fines." (114-115).

Las partes en la posición original, según podemos pensar, sufren exactamente de lo contrario de la incapacidad que aflige a casi todos nosotros, cuando afrontamos elecciones básicas en nuestra vida. En

general, sabemos lo que queremos, pero no sabemos cómo conseguirlo. Ellos, en cambio, saben cómo conseguir lo que quieren, pero no saben lo que quieren.

Como Rawls señala, la concepción de un plan racional de vida, en una o en otra versión, tiene una larga y honrosa herencia filosófica. En mi opinión, eso es fundamentalmente erróneo, por razones que apuntaré en la cuarta parte, pero, dada su progenie, Rawls difícilmente puede ser censurado por recurrir a ello, siempre que lo necesita. Y, como Rawls señala en la frase que precede al pasaje que acabamos de citar, la noción de un plan racional de vida "no es algo sobre lo cual discutan la doctrina contractualista y el utilitarismo" (114).

Pero el simple conocimiento de que se tiene un plan racional de vida no es todavía suficiente para proporcionar una base de elección racional entre principios de distribución, porque el solo hecho de que alguien tenga ese plan no le dice nada acerca de la relación de dicho plan con los tipos de cosas que se distribuyen. Por lo menos en algunos "planes de vida" (si pueden llamarse así) de base ultramundana, religiosa, nada de lo que al hombre le es posible distribuir puede tener el menor valor, incluidas la seguridad física y la vida misma. Después de todo, Jesús no dijo al joven rico: "Vete, vende todo lo que tienes y dáselo a los pobres, y tendrás un tesoro en el cielo: y ven, toma tu cruz y sígueme, *pero conserva lo necesario para poder realizar un plan racional de vida*". Para un verdadero cristiano, el único bien distribuable, digno de ser recibido, es la gracia, y la gracia viene de Dios, no de la sociedad.

Así, Rawls da por sentado que las circunstancias de la justicia incluyen una orientación secular (aun cuando, un poco extrañamente, parecen requerir, sobre todo, tolerancia religiosa), y formula su teoría de los *bienes primarios*. "Los bienes sociales primarios, presentados en amplias categorías, son derechos, libertades y oportunidades, así como ingresos y riquezas" (114). La clave de esta "sutil teoría" de los bienes primarios se pone de manifiesto inmediatamente, cuando Rawls nos dice que son los tipos de cosas de los que toda persona racional preferirá tener más que menos, cualesquiera que sean las particularidades de su plan de vida. Convenientemente registrados (porque son multidimensionales), los bienes primarios pueden ser tratados como una cantidad por la que todo agente racional tiene una utilidad marginal positiva.

Rawls establece, además, que sólo el registro de los bienes primarios debe tenerse en cuenta para la evaluación de la condición de un

hombre representativo en un sistema de instituciones, y ahora resulta clara la razón de estas maniobras. Cuando se tendió el velo de la ignorancia, las partes contratantes se vieron privadas de todas las bases para preferir un determinado conjunto de ordenamientos institucionales a otro. Al carecer de toda información acerca de sus intereses y de los medios necesarios para procurarlos, ni siquiera podían dar por sentado el tipo mínimo de utilidad marginal positiva generalizada (cualquiera que fuese la que estaba distribuyéndose) que justificase la adopción de un cuasi-ordenamiento de fuerte-preferencia-de-Pareto. De todo lo que podían conocer, en la vida real podían preferir menos a más. Pero, con un plan racional de vida, y con el conocimiento de que determinados bienes primarios son medios para algunos de esos planes, las partes se hallan, una vez más, en condiciones de afrontar el problema de la elección con alguna esperanza de resolverlo. Queda por ver si elegirían la versión menos aventajada del principio de diferencia en estas circunstancias, ahora un tanto complicadas, pero, por lo menos, puede darse al principio algún significado. En los términos que acabamos de introducir, se convierte en la orden de maximizar el registro de bienes primarios del hombre representativo menos aventajado de la sociedad.

Rawls pretende que el velo de la ignorancia hace soluble su problema. "El velo de la ignorancia hace posible la elección unánime de una determinada concepción de la justicia. Sin estos límites puestos al conocimiento, el problema de las negociaciones en la posición original sería infinitamente complicado" (167). Ahora está claro que la verdad es exactamente la contraria. La imposición del velo de la ignorancia hace el problema de las negociaciones infinitamente indeterminado, y, por consiguiente, totalmente insoluble. Aun para devolver la cuestión a la forma de un *problema*, y no simplemente de un encogimiento de hombros, Rawls tiene que atribuir un pleno conocimiento de los "hechos generales de la sociedad humana" a las partes, dar por sentado que todas ellas tienen planes racionales de vida, agregar una sólida teoría de bienes primarios, y, por último, establecer que las partes seleccionan sus principios de distribución, sin pensar más que en la distribución de bienes primarios. Cada una de esas adiciones al modelo es susceptible de fuertes objeciones, y constituye un signo de la debilidad del velo de la ignorancia, no de su fuerza, el hecho de que sean necesarias para que la teoría se sostenga en pie.

Además de los cambios impuestos a la teoría por las implicaciones del velo de la ignorancia, también hay ciertas clarificaciones, revisiones y elaboraciones exigidas por el nuevo principio de diferencia. Ante

todo, Rawls aclara explícitamente que el principio es de aplicación a la amplia y básica organización o disposición institucional de una sociedad, y no a cada equipo de béisbol, a cada club filatélico, y a cada tienda de comestibles.

Una sociedad, según Rawls, "es una asociación, más o menos auto-suficiente, de personas que reconocen ciertas reglas de conducta como obligatorias en sus relaciones, y que en su mayoría actúan de acuerdo con ellas. ... (A)demás... estas reglas especifican un sistema de cooperación diseñado para promover el bien de aquellos que toman parte en él" (20). No está muy claro cómo pretende Rawls que interpretemos esta definición. Para ver el problema que tal definición plantea, podemos recordar las observaciones que Max Weber formulaba a la tarea análoga de definir el estado. En los párrafos iniciales de su lección, "La política como vocación", Weber observaba:

"Sociológicamente, el estado no puede ser definido en términos de sus fines. Apenas hay alguna tarea de la que no se haya hecho cargo alguna asociación política, y no hay ninguna tarea de la que pueda decirse que ha sido siempre exclusiva y peculiar de aquellas asociaciones que se conocen como asociaciones políticas: hoy el estado, o, históricamente, aquellas asociaciones que han sido las predecesoras del estado moderno. Por último, puede definirse el estado moderno, sociológicamente, sólo en términos de los *medios* específicos que le son peculiares, como a toda asociación política, concretamente, el uso de la fuerza física."²³

La cuestión consiste en que, si definimos el estado en términos de su fin o propósito característico, nos veremos obligados a concluir que las asociaciones aparentemente políticas que no aciertan a alcanzar ese fin no son simplemente malos estados, sino que no son estados, en absoluto. En su definición de una sociedad, Rawls va más allá de la condición de diferenciación e integración funcionales de las actividades productivas y distributivas, para exigir que esas actividades estén regidas por normas que especifiquen un sistema de cooperación *diseñado* para promover el bien de aquellos que toman parte en él. La deducción natural que se extrae de esta definición es la de que el Sur anterior a la guerra, por ejemplo, no podía ser considerado como una *sociedad*, a no ser que, por lo menos, pudiera demostrarse que la esclavitud constituía un sistema de cooperación adecuado para promover los intereses

²³ From Max Weber: *Essays in Sociology*, trad., ed., y con una introducción de H. H. Gerth y C. Wright Mills (Oxford University Press, 1958), págs. 77-78.

de aquellos (incluidos los esclavos) que tomaban parte en él, y, más rigurosamente, que estaba *diseñado* para ese fin (no latente como opuesto a los fines manifiestos permitidos).

Ahora bien: en materia de definiciones, el autor es rey, de modo que Rawls es libre de utilizar de esa forma el término "sociedad". Pero como él, evidentemente, desea decir de muchas "sociedades" represivas que eran *sociedades injustas*, seguramente habría sido más juicioso el empleo de una definición que le permitiese tratar a esos grupos humanos como sociedades. La alternativa es la de insistir en que solamente los grupos humanos que satisfacen su definición son sociedades, y luego agregar, como una nueva premisa, el supuesto de que las partes en la posición original desean encontrarse en una "sociedad", cuando el velo de la ignorancia se levante, más que en cualquier otra organización humana.

Como Rawls conoce, indudablemente, las observaciones de Max Weber y el problema general de definición que nosotros hemos planteado, es natural suponer que tuvo alguna buena razón para construir una noción de promoción cooperativa del bien en su definición de la sociedad. Pero confieso que yo no la veo. Más aun: Rawls refuerza la impresión de que ha pretendido que su definición se neutralice, al introducir, inmediatamente, el concepto, más específico, de una "sociedad bien ordenada". Establece que se dirá que una sociedad está bien ordenada, "no sólo cuando está diseñada para promover el bien de sus miembros, sino cuando también está efectivamente regulada por una concepción de la justicia" (21).

A continuación, se explica que esto significa una sociedad "en la que: 1) cada cual acepta y sabe que los otros aceptan los mismos principios de justicia, y 2) las instituciones sociales básicas satisfacen generalmente estos principios y se sabe generalmente que lo hacen" (21).

En el nivel más elemental, el concepto de una sociedad bien ordenada es una generalización de una de las implicaciones procedimentales del juego de regateo original. En su forma primitiva, podemos recordar que los participantes en el juego eran personas racionales corrientes, plenamente conscientes de sus características individuales y de la sociedad en que vivían, y también (aunque Rawls no lo dice explícitamente) plenamente conscientes de las identidades y de las funciones sociales de los otros "jugadores". Las normas del juego del regateo garantizaban que cada jugador conocería, completa y conscientemente, la formulación precisa de cualquier conjunto de principios que surgiese del juego como solución de éste. Más aun: las condicio-

nes de racionalidad y de unanimidad garantizaban que cada jugador sabría que cada uno de los otros jugadores poseía la misma información. Por último, la condición del propio interés aseguraba que todo sistema de diferenciación e integración sociales que se ajustase a los principios elegidos por los jugadores sería aceptablemente descriptible como "un sistema de cooperación diseñado para promover el bien de aquellos que toman parte en él". Ahora que ha sido tendido el velo de la ignorancia, y destruido el carácter del juego del regateo de la posición original, Rawls tiene que introducir como estipulaciones lo que originalmente se seguía como estrictas implicaciones de las premisas del modelo.

A un nivel un poco más profundo, el supuesto de una sociedad bien ordenada tiene como finalidad la de permitir a Rawls la exclusión de ciertos tipos de versiones artificiosas, "de noble mentira", del utilitarismo. Por ejemplo, algunos defensores del principio de utilidad han sugerido que es realmente deseable, desde un enfoque utilitarista, para todos o para la mayoría de los miembros de una sociedad, mantener creencias no utilitaristas acerca de la justicia, creencias cuyo rígido compromiso con un principio inflexible es incompatible con la posibilidad de ejemplo utilitarista de la normas generales. Esos pensadores podrían incluso estar de acuerdo en que convendría al propósito de maximización de la utilidad el hecho de que la teoría de Rawls alcanzase una aceptación universal, aunque fuese falsa (insistirían privadamente) como teoría de la justicia. A esto, Rawls replica que la exigencia de un buen ordenamiento excluye cualquier doblez del tipo de Gran Inquisidor (502, nota).

Yo creo que, en el nivel más profundo, la definición idealista de la sociedad, como podemos denominarla, y la ulterior estipulación de un buen ordenamiento expresan el gran compromiso de Rawls con la visión de una sociedad estable en la que justicia y bondad sean congruentes.

Con toda la sociedad como objeto de investigación, Rawls toma la "estructura básica" de esa sociedad como el patrón o conjunto de disposiciones al que ha de aplicarse el principio de diferencia. "O, más exactamente —dice—, el modo en que las instituciones sociales más importantes distribuyen los derechos y deberes fundamentales y determinan la división de las ventajas provenientes de la cooperación social" (23). Como anticipo de su diferenciación entre dispositivos políticos y económicos, que nosotros no agregaremos a este análisis hasta la sección siguiente, Rawls identifica las instituciones más importantes co-

mo la constitución política y los principales dispositivos sociales y económicos. Después, como ejemplos de esas instituciones más importantes, cita "la protección jurídica de la libertad de pensamiento y de conciencia, la competencia mercantil, la propiedad privada de los medios de producción y la familia monogámica" (23)²⁴.

Más adelante, sin embargo, en la sección titulada "Las posiciones sociales relevantes", Rawls simplifica considerablemente su análisis, estableciendo que "esencialmente, cada persona tiene dos posiciones relevantes: la de igual ciudadanía y la definida por el lugar que ocupa en la distribución de ingresos y de riqueza" (118). Una vez que ha introducido la norma de la prioridad de la libertad, sus dos principios exigirán, simplemente, que todos tengan garantizada una segura condición de miembros de estas funciones sociales, y que, sometida a esta exigencia, se maximice la expectativa del ocupante representativo del "último lugar en la distribución de ingresos y de riquezas". Rawls difícilmente puede contribuir a hacer esta simplificación, dadas las tareas lógicas que a sí mismo se ha fijado y habida cuenta del tipo de teoría social que parece dispuesto a invocar en su ayuda. Si hubiera de abordar seriamente la noción de que instituciones sociales tan dispares como el matrimonio monogámico y la propiedad privada de los medios de producción han de ser tratadas como "instituciones sociales más importantes", tendría un imposible problema de orden unido a un insoluble problema de índice. El problema de orden surgiría porque cada individuo ocupa lugares en un cierto número de diferentes instituciones sociales importantes, y no estaría claro cuál de ellos se elige como el lugar que determinará su categoría en la jerarquía social. El problema del índice surgiría cuando se tratase de comparar la condición de la mujer representativa en la institución del matrimonio monogámico con la del obrero representativo no calificado en la institución de la producción industrial. Pero, aunque Rawls, en realidad, no tenía elección, no logró evitar algunas objeciones muy serias, como veremos en la cuarta parte.

Por último, tendido el velo de la ignorancia, asegurado un conocimiento de los hechos básicos, localizados los planes de vida, con los bienes primarios como objetivo, con las instituciones sociales más importantes como tema de deliberación, con la exigencia de una sociedad

²⁴ No está apoyando la propiedad privada de los medios de producción ni la familia monogámica, aunque se tiene la sensación de que aprueba la una y la otra. Simplemente, está presentándolas como ejemplos de instituciones importantes.

bien ordenada, y con la elección reducida a los principios que especifican diferentes normas para definir las atribuciones a los lugares que se ocupen en la distribución de ingresos y riqueza, Rawls todavía tiene que decir algo convincente acerca de los principios de elección racional a los que habrán de apelar las partes en la posición original. Es en este punto del desarrollo de la teoría donde se introduce una defensa formal de la regla maximin de elección bajo incertidumbre²¹.

El argumento de Rawls en favor de la regla maximin es complejo, y su categoría lógica no está enteramente clara. Hay momentos en que parece asegurar que se trata de una regla que unos agentes racionales, en cuanto tales, considerarían prudente adoptar, si la deliberación se efectuase desde un punto de vista "kantiano" (después volveremos sobre este concepto); pero, en otros momentos, dice que ha reajustado su caracterización de la posición original, a fin de asegurar que en ella se elegirá la regla maximin. En la cuarta parte, daremos al tema un tratamiento detallado. Aquí, baste resumir brevemente las tres razones de Rawls para pretender que las partes en su posición original optarían por la maximin. Señala, primero, que la regla es apropiada para los casos en que "resulte imposible el conocimiento de las alternativas posibles, o que al menos sea altamente inseguro" (183). Segundo, es una regla apropiada para los casos en que el razonador concede muy poca importancia a los incrementos adicionales que podría obtener por encima del mínimo que puede garantizarse a sí mismo maximizando ese mínimo. Y, tercero, es una regla apropiada para los casos en que la amenaza de pérdidas por debajo de ese mínimo garantizado tiene una gran importancia en la valoración del razonador. El velo de la ignorancia se propone hacer imposibles los cálculos de probabilidades (en especial, porque tales cálculos tendrían que combinar estimaciones y evaluaciones de todas las prácticas alternativas posibles con independientes estimaciones de la posibilidad de acabar en cada función representativa de cada una de esas prácticas). La teoría de los bienes primarios, juntamente con el principio de la prioridad de la libertad (todavía no introducido en nuestro análisis), reduce, en opinión de Rawls, el

²¹ Una vez más, me permito recordar al lector que este ensayo es una reconstrucción del desarrollo filosófico de Rawls, no una historia del mismo. Según creo, los argumentos acerca de la maximin figuraron entre los primeros que integraron el pensamiento de Rawls, hace un cuarto de siglo. Mi propósito al introducirlos en este punto es el de mostrar la función lógica de los mismos en su teoría y el de aclarar cuáles son los problemas que los requieren aquí, más que en cualquier otra fase anterior.

valor de los incrementos que pueden obtenerse por encima del mínimo. Y se supone que el peligro de una insuficiencia real de bienes primarios convierte en inaceptable el riesgo de resultados inferiores a ese mínimo garantizado por el principio de diferencia.

Todos los rasgos importantes de la teoría final de Rawls se encuentran ahora localizados, excepto el de la prioridad de la libertad. (Las estipulaciones relativas a la tasa de ahorro y a otras materias similares no son, en mi opinión, fundamentales para la teoría, aunque son, ciertamente, problemas importantes para la filosofía social). No estamos preparados todavía para someter esta teoría a una crítica sistemática, pero casi hemos llegado al fin del proceso de desarrollo. Con algunas excepciones, las objeciones que ahora resultan oportunas son objeciones a la teoría de Rawls en cuanto tal, y no problemas de las versiones anteriores que condujeron a Rawls a alterar o a reajustar la teoría. Hay, sin embargo, una réplica colectiva que Rawls puede hacer, y ciertamente hace, a las objeciones del género que yo formularé: concretamente, la de que no alcanzan a la íntegra totalidad de la teoría. Esta cuestión de la totalidad, y la licitud de invocarla como justificación de elementos particulares de la teoría, atañe a la categoría lógica de las aseveraciones hechas a lo largo del libro. Pospongo mi discusión de la misma hasta cerca del final de este ensayo, momento en que atenderé precisamente a la noción de Rawls del "equilibrio reflexivo".

IX

LA PRIORIDAD DE LA LIBERTAD Y OTRAS COMPLICACIONES

COMO SABEN los lectores de Rawls, hay todavía una enorme cantidad de ulteriores matices que agregar, antes de que nos hallemos en presencia de la teoría completa. Prescindiré de la discusión de la mayor parte de ese material, por motivos que han sido indicados ya. Pero hay algunas materias que requieren cierto tratamiento o que plantean interesantes problemas. La más importante, sin duda, es la pretensión de Rawls de que las partes en la posición original decidirían hacer el primer principio, relativo a la libertad igual, lexicográficamente anterior al segundo principio, que especifica las reglas de distribución. En esta sección, se tratarán la prioridad de la libertad y algunas cuestiones menores. En la tercera parte, que sigue a esta sección, exploraremos la relación entre Kant y Rawls, de un modo que esclarecerá más aun —así lo espero— la relación de la libertad con la distribución.

LA PRIORIDAD DE LA LIBERTAD²⁶

En el modelo original del juego del regateo, como hemos visto, la meta de las deliberaciones es un conjunto de principios para la evaluación de unas prácticas particulares. Aunque Rawls no lo dice, e incluso dice algunas cosas aparentemente contrarias, se comprende claramente que ciertas condiciones fundamentales —en general, las que podrían llamarse condiciones de democracia constitucional—, o bien se presuponen, o bien se crean en la estructura del propio juego. De igual modo que se considera que el libre comercio, en la teoría económica clásica, descansa en una base de libertades políticas y de protecciones legales que garantizan el derecho de cada persona a establecer contratos, a

²⁶ Muchas de mis observaciones sobre la prioridad de la libertad siguen muy estrechamente la línea de argumentación desarrollada por H. L. A. Hart, en su ensayo, "Rawls on Liberty and its Priority", publicado inicialmente en *University of Chicago Law Review*, Vol. 40, núm. 3 (Primavera, 1973), págs. 534-55, y reimpresso en *Reading Rawls*, editado por Norman Daniels. Ver especialmente Sección III, "Limiting Liberty for the Sake of Liberty".

exigir el cumplimiento de los mismos o una compensación por su incumplimiento, etc., así también la presentación que Rawls hace de su modelo me parece que propone, inequívocamente, un cierto sistema similar de derechos políticos y de protecciones como marco del juego del regateo.

De todos modos, ya desde el principio mismo, Rawls es, por lo menos, ambivalente en cuanto a la relación de su principio económico de justificada desigualdad de distribución con los principios políticos e instituciones de ciudadanía igual. Esta ambivalencia aparece, como he señalado, en la primera formulación del primer principio de la justicia, que habla de "la más amplia libertad", aun cuando el contexto de la exposición aclara que Rawls se refiere "al mayor volumen de bienes y servicios".

Con el cambio de enfoque desde las prácticas micro-económicas a las instituciones macro-sociales, y con el velo de la ignorancia tendido, Rawls reconstruye sus dos principios, de tal modo que el primero se convierte en un principio político de ciudadanía igual, y el segundo incorpora la noción completa de unas justificables desviaciones a partir de una distribución económica igual. Hablando histórica y extrasistemáticamente, podríamos decir que, en la forma revisada de los dos principios, el Principio I enuncia la esencia del sistema de igualdad legal y política que se desarrolló a finales del siglo XVIII y en el XIX como plataforma para las libres operaciones del capitalismo industrial, y el Principio II define las normas de justicia social que han de emplearse para atenuar las desigualdades y las injusticias de aquellas operaciones.

Cuando Rawls pasa a la teoría de los bienes primarios, enumera "derechos, libertades y oportunidades, así como ingresos y riquezas" (114). En lo que él llama la concepción más sencilla de la justicia, tenemos que imaginar un índice de este heterogéneo conjunto de "bienes", de modo que podemos establecer, por lo menos, una relación ordinal de conjuntos alternativos y después maximizar el índice asignado al hombre representativo menos aventajado²⁷. En consecuencia, divi-

²⁷ La teoría es bastante problemática, sin enredarse en distinciones entre medidas ordinales y cardinales de bienes primarios, pero recordemos que, sólo con una relación ordinal, ni siquiera pueden abordarse cálculos de un supuesto valor del índice en situaciones arriesgadas. En vista de la ignorancia de las partes en la posición original, podría pensarse que tales consideraciones se excluían, pero las partes podrían desear hacer cálculos utilizando probabilidades condicionales, basadas en su conocimiento de los hechos básicos de la sociedad humana. También podrían desear hacer

de el conjunto de bienes primarios en dos subconjuntos: derechos y libertades, y lo demás. Entonces, impone una regla de prioridad a los Principios I y II, cuyo efecto es el de requerirnos para que maximicemos el índice de derechos y libertades antes de dedicar nuestra atención a la maximización de oportunidades y de poderes, de ingresos y de riquezas. En la última revisión de los principios, Rawls admite una sociedad en las primeras fases de su desarrollo social y económico, en la que se cancele la norma de prioridad y se sacrifique un cierto grado de libertad por una suficiente mejora en el bienestar material, pero sólo a fin de avanzar hacia el momento en que pueda imponerse todo el rigor de la norma de prioridad. El efecto claro de la norma de prioridad, como Rawls explica, es el de permitir limitaciones en la libertad, sólo por causa de la libertad misma.

Es bastante fácil comprender por qué Rawls procedió a la versión revisada de sus principios de justicia, a pesar de las complejidades que la revisión introduce en la teoría. En primer lugar, la prioridad de la libertad articula la convicción de Rawls, evidentemente profunda, de que el mutuo respeto de la ciudadanía igual expresa el reconocimiento recíproco de la personalidad moral por parte de los hombres. Es, como veremos, la forma en que Rawls incorpora a su teoría el imperativo kantiano de tratar a la humanidad, tanto en sí mismo como en los otros, siempre como un fin y nunca simplemente como un medio. Malvender una porción de las propias libertades por una vida más cómoda equivaldría, en opinión de Rawls, a vender la primogenitura como ser humano por un plato de lentejas.

Al propio tiempo, el uso que Rawls hace de modelos cuasi-económicos de razonamiento dificulta extremadamente la acumulación de consideraciones políticas y económicas en un solo modelo o "índice". Mediante la partición del conjunto de bienes primarios y la imposición de la regla de prioridad, Rawls puede separar las cuestiones esencialmente políticas de estados legales, garantías constitucionales y formas de gobierno, de las cuestiones económicas de la estructura del mundo del trabajo, de las actividades del mercado y de la distribución de ingresos y posiciones sociales. Es por esta razón, entre otras, por lo que Rawls atribuye precisamente dos posiciones importantes a cada perso-

esos cálculos en la segunda, tercera o cuarta etapa del levantamiento del velo. Un índice ordinal haría imposible la elección con riesgo, en cualquier etapa. Cuando discutamos los argumentos de Rawls en favor de la maximin, veremos que presuponen funciones de utilidad cardinal.

na en la estructura básica de una sociedad. La primera —ciudadanía igual— corresponde al principio de libertad igual; la segunda —el lugar que una persona ocupa en la distribución de los ingresos y de la riqueza— corresponde al principio de diferencia.

Por último, el requisito de que la regla de prioridad entre en juego sólo por encima de un decoroso nivel mínimo de bienestar es necesario para excluir juicios manifiestamente absurdos. Si se observase una prioridad estricta, ni siquiera un descenso catastrófico en el índice de bienes primarios, suficientemente grave para amenazar con la extinción de la sociedad, podría justificar ni la más ligera relajación de la libertad igual, formal, constitucional, con la que la sociedad avanzaba hacia su destrucción, aunque pudiera haber muchas razones para suponer que sería posible comprar la supervivencia, al precio de una tiranía temporal.

¿Qué tenemos que hacer con el principio de la prioridad de la libertad? Dejando a un lado la cuestión de su verdad última y también toda consideración acerca de su conformidad con las intuiciones morales de una u otra persona, formulemos sencillamente dos preguntas: primera, ¿tiene sentido el principio, podemos comprender lo que significa y cómo debería aplicarse?; y, segunda, ¿lo elegirían las partes en la posición original?

Poco después de iniciada la segunda mitad de *Teoría de la justicia*, Rawls acaba por establecer sus dos principios en su forma final, ya completada con las reglas de prioridad. He aquí la parte que interesa a nuestra discusión:

Primer Principio

Cada persona ha de tener un derecho igual al más amplio sistema total de libertades básicas, compatible con un sistema similar de libertad para todos.

Primera Norma de Prioridad (La Prioridad de la Libertad)

Los principios de la justicia han de ser clasificados en un orden lexicográfico, y, por tanto, las libertades básicas sólo pueden ser restringidas en favor de la libertad en sí misma. Hay dos casos:

- a) una libertad menos extensa debe reforzar el sistema total de libertades compartido por todos;
- b) una libertad menor que la libertad igual debe ser aceptada por aquellos que tienen una libertad menor (340-341).

Rawls ha simplificado, evidentemente, su problema de clasificación, separando los derechos y las libertades, de las oportunidades y

los poderes, los ingresos y la riqueza, en la clase de bienes primarios. Pero como se adhiere a un modelo cuasi-económico de maximización, todavía se ve obligado por la forma de su principio a recurrir a algún procedimiento de clasificación para las libertades. Si estuviesen elaborándose, simplemente, modelos formales, un primer paso podría consistir en un análisis multidimensional de la libertad, y la invocación de una cuasi-exigencia de unanimidad. Todo cambio en las libertades básicas de un individuo que no disminuyese su libertad en ninguna dimensión y que incrementase su libertad, por lo menos, en una dimensión, sería un cambio positivo; y así, sucesivamente. Entonces, habría que afrontar los problemas de agregación, o de clasificación, a fin de llegar a un ordenamiento completo, suficiente para decir si algún sistema total de libertades básicas era más extenso, menos extenso o igualmente extenso que cualquier otro sistema total. Como H. L. A. Hart arguye, muy convincentemente, en el ensayo citado al comienzo de esta sección, esas clasificaciones y comparaciones dentro del plano de la libertad política y legal están muy lejos de ser fáciles. Muchos de los más profundos e insuperables desacuerdos políticos entre ciudadanos conscientes y honorables toman, precisamente, la forma de juicios diametralmente opuestos sobre la aceptabilidad de limitar una libertad por amor de otra.

Pero, en mi opinión, hay un problema todavía más profundo, derivado de la procedencia económica del principio de libertad igual de Rawls. Yo puedo comprender, tanto histórica como institucionalmente, lo que significa decir que, en un sistema político, todos los hombres son iguales ante la ley. Esa frase suscita la exigencia de que todas las personas sean responsables ante un solo tribunal, que no haya tribunales eclesiásticos para el clero, ni una Cámara de los Lores para los pares. Puedo comprender también el principio de igualdad política, que se traduce en el sufragio universal, en el derecho universal a ocupar un puesto político, etc. Pero me resulta muy difícil formular una noción utilizable de una *medida* de la libertad, que me permita decidir si un determinado sistema de leyes consiente un mayor o menor volumen, por así decirlo, de libertad de expresión.

Aun a riesgo de parecer jocoso, permítaseme formular algunas preguntas sencillas, incluso ingenuas. ¿Mayor libertad de expresión significa libertad para hablar más frecuentemente, o a más gente, o acerca de cuestiones más importantes? ¿Significa libertad para apartarse más del canon de opiniones permitidas socialmente aceptado? La simple formulación de preguntas de este carácter revela con toda claridad que

se está enfocando el tema de un modo equivocado. Y, sin embargo, no veo cómo Rawls puede dejar de someterse a tales indagaciones, una vez que adopta la fórmula de "la libertad más amplia, compatible con una libertad igual para todos".

Para objeciones de este carácter, Rawls tiene dos respuestas. La primera, que es más bien elegante, consiste en la distinción entre la libertad y el valor de la libertad (237). Esto le permite rehuir el frecuente sarcasmo de que, en una democracia capitalista, el pobre y el rico tienen el mismo derecho legal a comer en un restaurante caro. Pero el problema de clasificación no se plantea porque unas libertades iguales sean de valor desigual para personas de diferentes niveles en la jerarquía de oportunidades y poderes, de ingresos y riqueza. Se plantea, a causa de la necesidad de ordenar diferentes grados de libertad y agregar luego unos cuantos de libertad de diferentes tipos en una sola medida de la magnitud de un sistema total de libertad. La verdadera respuesta de Rawls al problema es, sencillamente, la de retirarse a sus intuiciones morales, recordando al lector que hay límites a la precisión, que pueden alcanzarse mediante alguna teoría moral. Pero una retirada a la intuición es una maniobra aceptable, sólo a condición de que establezcamos claramente las fronteras de una teoría moral. Como Hart señala, sin embargo, nuestras incertidumbres se presentan en el corazón mismo del campo de la libertad política. Rawls, simplemente, no explica lo que *significa* su principio, cuando éste habla de "la más amplia libertad". La respuesta a nuestra primera pregunta es, por lo tanto, no: no podemos comprender lo que significa el principio, ni cómo debería ser aplicado.

¿Elegirían las partes en la posición original el primer principio, como parte de su solución al problema de elección racional con que se enfrentan? ¿Y le atribuirían la norma de prioridad extremadamente fuerte enunciada por Rawls? Según la interpretación de la posición original tan decididamente apuntada por las primeras formulaciones de la teoría de Rawls, y expuesta en el lenguaje y en el razonamiento de *Teoría de la justicia*, la respuesta me parece que debe ser un *no* cualificado. Pero la prioridad de la libertad igual puede defenderse más sólidamente desde la interpretación kantiana de la posición original, y sospecho que éste es el motivo de que la sección acerca de la interpretación kantiana siga inmediatamente a la sección titulada "Consideraciones sobre la prioridad de la libertad".

El origen de la dificultad es la idea de un plan de vida, según Rawls la utiliza para dar una estructura al problema de la elección en

la posición original. *Teoría de la justicia*, en cuanto manifiesto de una fe política, se encuentra en la gran tradición que se remonta hasta la *Política* de Aristóteles de la celebración de la *polis*, de la comunidad de agentes racionales que encuentran su más alto bien y su más completa realización en el discurso racional y en la deliberación colectiva mediante los cuales gobiernan sus asuntos comunes. Es esta visión de la sociedad justa como fin de la vida, y no sólo como medio, la que impulsa a Rousseau a escribir:

"Cuanto mejor constituido se halla un estado, más intervienen los asuntos públicos en los asuntos privados, en los espíritus de los ciudadanos. Las cuestiones privadas incluso llegan a ser considerablemente menos, porque cada individuo participa tan ampliamente en la felicidad común, que no tiene tanta necesidad de procurarla en los recursos privados."²⁸

Yo también encuentro sugestiva esta visión de la sociedad justa, aunque observo una considerable ambivalencia en cuanto a la pérdida de la condición privada y a la intrusión que ella implica. Y una parte en la posición original que supiese que tal visión informaba su plan de vida optaría, muy probablemente, por un principio rawlsiano de la más amplia libertad igual, si podía explicárselo, completado con una forma fuerte de la regla de prioridad. Pero hay otros planes de vida, igualmente racionales, aunque no según la concepción más cargada de racionalidad, en los que la libertad política es, simplemente, un medio para fines privados, y no uno de los principales fines en sí misma. Para personas con tales planes de vida, las relaciones entre la libertad y los ingresos serían muy interesantes, al menos dentro de los límites evidentemente amplios destinados a protegerse frente a las más graves restricciones de los privilegios. Sabiendo solamente que tenían planes racionales de vida, e incluso sabiendo que la ejecución afortunada de tales planes exigiría bienes primarios (cuantos más, mejor), las partes en la posición original podrían no tener modo alguno de saber si sus planes de vida bastarían para justificar un fuerte hincapié en la libertad como opuesta a los ingresos, ni razón alguna, en absoluto, para optar por un ordenamiento lexicográfico absoluto de las dos subclases de bienes primarios.

Para establecer tan sólida conclusión, Rawls tiene que encontrar alguna forma de demostrar que el simple hecho de ser un agente moral racional requiere que las libertades de un orden político democrático se sitúen absolutamente por encima de los ingresos y de la riqueza, al

²⁸ *El contrato social*, Libro III, capítulo XV.

menos una vez que se haya asegurado un mínimo decoroso. Rawls descubre ese argumento en la interpretación kantiana de la posición original.

OTRAS COMPLICACIONES

Aun después de hacer todos los reajustes al modelo de las fases precedentes, Rawls comprueba que todavía no tiene los materiales para un teorema, y entonces introduce unas pocas complicaciones adicionales. Como el propósito de este ensayo es el de poner de manifiesto la estructura subyacente del desarrollo de la teoría de Rawls, más que comentarla *seriatim*, no trataré de seguir todas las vueltas y revueltas de la exposición. Sin embargo, hay unos pocos puntos que requieren algún comentario.

En realidad, por lo que se refiere al problema de la elección racional tal como se plantea en la posición original, Rawls debe definir un espacio de regateo de posibles puntos de solución y luego presentar argumentos para demostrar que los jugadores se pondrían de acuerdo sobre un determinado punto o conjunto de puntos de ese espacio. En la forma original del modelo, las "soluciones" eran principios de distribución de un conjunto de mercancías, y la homogeneidad del problema parecía suscitar la esperanza de que pudiera definirse tal espacio de solución. Pero, en la versión final, las alternativas entre las cuales tienen que elegir las partes, son, en realidad, nada menos que la totalidad de las filosofías sociales, y no hay una forma clara de identificar las dimensiones a lo largo de las cuales esas filosofías pueden cambiar, de modo que definan un espacio de soluciones utilizables. En consecuencia, Rawls sustituye la versión del problema de elección inicialmente ideada, por otra versión simplificada, estipulando que las partes contarán con una relación limitada de alternativas para elegir entre ellas. (Ver Sección 21.) Las alternativas presentadas a las partes en la posición original son, esencialmente, diversas formas de utilitarismo, ciertas doctrinas intuicionistas, los dos principios de la justicia del propio Rawls, y algunas de las más aceptables mezclas de estos elementos.

Rawls tiene clara conciencia del carácter insatisfactorio de esta simplificación. Como él dice: "admito que es un modo poco satisfactorio de proceder" (148). Pero considero justo decir que los inconvenientes del procedimiento no le preocupan mucho, porque, según he apuntado en mis observaciones iniciales, Rawls está realmente entregado a una

disputa con el intuicionismo y con el utilitarismo. No hay nada insólito ni censurable en el hecho de que un filósofo concentre su atención en los que él cree que son sus principales competidores filosóficos e ignore otras alternativas, pero tal reducción del enfoque erosiona considerablemente la pretensión de estar examinando las cuestiones *sub specie aeternitatis*.

Hemos señalado ya la restricción sobre la prioridad de la libertad —que producirá su efecto, sólo después de que en la sociedad se haya alcanzado un nivel suficiente de bienestar material. Rawls añade un nuevo matiz a su declaración general de los dos principios, en el sentido de que la sociedad como conjunto adoptará una tasa justa de ahorros.

A pesar de la elaboración económica de su discusión, el tratamiento que Rawls da al tema de una adecuada tasa de ahorros difícilmente puede calificarse de éxito. El problema no reside tanto en sus conclusiones como en el tema mismo. Sencillamente, no parece que allí se encuentre ninguna respuesta razonable a la pregunta: ¿cuánto debe sacrificar esta generación para el bienestar de las generaciones futuras? La pregunta, como se sabe, es bastante real. Desde luego, en países que combinan un bajo nivel de desarrollo económico con un alto grado de centralización del poder de decisión es, probablemente, la pregunta más importante que el gobierno puede plantearse a sí mismo y a sus consejeros económicos. Pero, si puedo introducir mis intuiciones morales en el análisis, descubro, tras la debida reflexión, que ni siquiera me es posible llegar a una elección aproximativa de una respuesta centrándome en consideraciones de justicia.

Ahora bien: cuando un problema de importancia fundamental de filosofía moral o social parece resistirse enteramente a un enfoque desde un punto de vista particular, puede ser un indicio de que el punto de vista es totalmente erróneo. Me gustaría esbozar unas pocas observaciones, con el fin de sugerir que el problema de una tasa justa de ahorros revela una profundísima inadecuación en la orientación filosófica básica de Rawls, y también —podría añadir— de las tradiciones utilitarista, intuicionista y del contrato social.

La dificultad particular de elegir una tasa de ahorros surge del hecho de que las partes no saben de qué generación de su propia sociedad proceden. Recuérdese que, en el modelo de Rawls, la tarea de las partes no es la de seleccionar una tasa *justa* de ahorros, sino solamente la de elegir una tasa de ahorros que sea racionalmente egoísta, bajo las peculiares restricciones de conocimiento de su situación. Se supone que

todo el quid de la posición original consiste en que, en virtud de la pura justicia procedimental de la situación, cualquier elección que las partes realicen *será* justa. No ahorrar nada, ni siquiera para la generación inmediata, parece claramente imprudente; ahorrar todo lo que se pueda, a lo largo de muchas generaciones, a fin de que las últimas generaciones de la sociedad puedan vivir en una comodidad total, parece igualmente imprudente. Y optar por un término medio entre estos dos extremos tiene toda la apariencia de una solución *ad hoc* de un dilema insoluble. Naturalmente, consideraciones de eficacia y de productividad marginal descendente pueden reducir un poco el alcance de las posibles alternativas, pero no hasta el punto de transformar el problema en una cuestión de elección racional.

Si tratamos de imaginar versiones reales del dilema, podemos ver por qué en el corazón del problema no se encuentra la justicia ni la elección racional. ¿Debería el pueblo chino sacrificar una generación, efectivamente, mediante las más rigurosas restricciones soportables en el consumo, a fin de elevar al máximo la tasa de acumulación de capital, en beneficio de las futuras generaciones? ¿O deberían adoptar una tasa de crecimiento más lenta, de modo que quienes ahora viven y trabajan puedan gozar de algunos de los frutos de su trabajo, antes de que sean demasiado viejos? Esta no es una pregunta para la que haya una respuesta moral correcta. Podemos estar de acuerdo en que su elección, cualquiera que sea, debe ser suya y no de sus gobernantes. Pero, si como pueblo decidiesen libremente sacrificarse por las generaciones futuras, ¿podríamos decir, con razón, que hubieran obrado injustamente? Y si, por el contrario, decidiesen gozar de una modesta elevación en su nivel de vida, incrementando más lentamente sus instalaciones industriales, ¿podríamos, con razón, calificar *esa* elección de injusta?

El origen de esta dificultad es la forma en que Rawls nos habría presentado la cuestión. Rawls concibe el punto de vista moral como una superioridad intemporal, desde la que contemplamos todo el tiempo y todo el espacio, ecuánime e isotrópicamente, como Lucrecio contemplaba el campo de batalla. Así, la elección de una tasa de ahorros, hecha por una generación para todas las generaciones, es formalmente idéntica a una elección hecha por un conjunto adecuado de todos los agentes racionales, en un momento dado, para todos los agentes racionales de ese momento. El hecho de que los no incluidos en la elección lleguen en un tiempo ulterior (o hayan llegado en un tiempo anterior) no debe tratarse de un modo distinto del hecho de que los no

incluidos en algún acto de elección se encuentren situados en un lugar diferente de los que ocupan quienes están incluidos.

Pero la existencia humana no es accidentalmente temporal; es esencialmente temporal. Lo que convierte en una cuestión de *justicia* la forma en que un subgrupo decide por toda la sociedad es el hecho de que, en principio, todo el grupo *podría* ser incluido en la decisión. Lo que convierte, al parecer, en una cuestión de justicia la forma en que los padres deciden por los hijos es el hecho humano de que las generaciones se superponen, de modo que los hijos, los padres y los abuelos tienen que vivir, durante un tiempo, en el mismo mundo. Lo que manifiestamente *no* convierte en una cuestión de justicia la forma en que esta generación decide por una generación lejana en el futuro es la certidumbre de que ambas no pueden compartir el mismo mundo, y, por consiguiente, tampoco podrían, en principio, reunirse para compartir el acto de decidir.

El velo de la ignorancia crea una situación de elección en la que las características *esenciales* de la existencia humana se dejan a un lado, juntamente con los accidentes de las variaciones individuales. En mi opinión, el resultado no es un punto de vista moral, sino un punto de vista no humano, desde cuya perspectiva las cuestiones morales no se clarifican, sino que se deforman y se distorsionan.

Es muy poco lo que hay que decir acerca de detalles como el levantamiento en cuatro etapas del velo de la ignorancia. Los filósofos sistemáticos tienen el hábito de descarnar los esqueletos de sus teorías con elaboraciones cuya relación lógica con la propia teoría podría describirse como una relación de posibilidad, más que de actualidad o necesidad. Cuando Hobbes permuta y combina los elementos de su teoría del egoísmo psicológico para generar una serie de brillantes definiciones de predicados de carácter, cuando Kant acierta a introducir en las casillas vacías de su arquitectura su particular y dinámica teoría de la materia, no debemos imaginar que éstas sean deducciones de las primeras premisas del *Leviatán* o de la *Crítica*. Ni siquiera podemos suponer que las teorías centrales de esas obras presten ningún crédito especial a los materiales subsidiarios que las han completado. Más bien, Hobbes y Kant (y también otros innumerables filósofos sistemáticos) demuestran, sencillamente, que sus teorías ofrecen un espacio, admiten, pueden encontrar una clara zona para una cierta psicologización, o moralización, o especulación científica, a las que ellos son muy adictos.

Creo que podemos adoptar una visión similar de una buena parte

X

KANT Y RAWLS

CUANDO RAWLS tiende el velo de la ignorancia sobre los participantes en el juego del regateo en la posición original, transforma tan completamente la situación de los mismos, que ésta deja de ser, en todo sentido reconocible, un juego del regateo. Rawls asegura, como si ello fuese una de las virtudes del velo de la ignorancia, que, bajo él, "las partes no tienen base para regatear, en el sentido usual del término" (166). Por razones que desarrollaré en la cuarta parte, me parece que este cambio socava todo el edificio de la teoría de Rawls y destruye lo que era la idea central de la primera forma del modelo. Un tanto sorprendentemente, Rawls acepta la transformación y la defiende como una de las bases de la plausibilidad moral general de su teoría.

Como hemos visto, la introducción del velo de la ignorancia era necesaria por la sencilla indemostrabilidad del teorema en su forma original. Pero el atractivo del velo de la ignorancia procede bastante claramente del hecho de que enlaza la línea de análisis de Rawls con una tradición totalmente diferente de filosofía moral, a la que él se siente atraído tan fuertemente como a la tradición económica clásica anglo-americana de la que proceden sus modelos de análisis y de argumentación: concretamente, la filosofía moral de Immanuel Kant.

Rawls ofrece una "interpretación kantiana" explícita de la posición original, en su versión final del velo de la ignorancia, y en breve examinaremos esa interpretación. Pero vale la pena señalar las otras dos formas en que Rawls parece repetir la táctica y el estilo filosófico de Kant, la primera de ellas muy profunda, y la segunda simplemente superficial. Atendamos primero a la profundidad²⁹.

Los primeros trabajos filosóficos de Kant se centraron en sus intentos de mediar en el conflicto teórico entre la escuela científica inglesa de Newton y sus seguidores. Sus escritos iniciales se inclinaban prime-

²⁹ En el resto de la tercera parte, me referiré frecuentemente a mis anteriores estudios de la filosofía de Kant. Ver especialmente *Kant's Theory of Mental Activity* (Harvard University Press, 1963) y *The Autonomy of Reason: A Commentary on Kant's Groundwork of the Metaphysics of Morals* (Harper & Row, 1973).

ro a un lado y luego al otro de la famosa disputa, pero, en 1770, Kant encontró lo que él consideró que era una solución satisfactoria del conflicto. En la *Disertación Inaugural* presentada con motivo de su elevación a la cátedra de la Universidad de Königsberg, Kant proponía, en realidad, una transacción o fusión de las dos escuelas. Transformando lo que había sido una discusión metafísica y científica en una cuestión epistemológica, Kant sostenía que ni la razón ni los sentidos constituían la única fuente de nuestro conocimiento de los objetos. Afirmaba que la razón nos da el conocimiento de las cosas tal como ellas realmente son, por medio de conceptos puros, nacidos de los recursos interiores de la razón; y los sentidos nos dan el conocimiento de las cosas tal como se nos aparecen, bajo las formas de espacio y tiempo. En la *Crítica de la razón pura*, Kant abandonó la pretensión de que la razón pueda facilitar un conocimiento de las cosas tal como éstas son en sí mismas, pero continuó describiendo la sensibilidad como una limitación del conocimiento. Decía que nuestras facultades de razón teórica, por sí solas, no podrían darnos nada más que las vacías tautologías de la lógica. Pero, si a la forma pura de la razón teórica agregamos la restricción de la sensibilidad —es decir, si nos limitamos a afirmaciones condicionalmente *a priori*, relativas a objetos de una experiencia posible (sensible) en general—, entonces podremos llegar a conclusiones sumamente sólidas, perfectamente generales. Podríamos, en efecto, establecer la validez (condicionalmente) *a priori* de la física pura y de la matemática pura. Aunque nuestras conclusiones se verían limitadas por las restricciones de las condiciones del conocimiento humano, serían perfectamente generales dentro de esa amplia esfera, y se establecerían enteramente *a priori*, sin recurrir a la observación o a la experiencia particular.

Quisiera indicar que la forma original de la teoría de Rawls se levanta sobre una maniobra que presenta una sorprendente semejanza con esta idea de Kant. En efecto, podemos imaginar a Rawls diciendo que los principios puramente formales de la razón práctica, tal como se explican en la teoría de la elección racional individual, no bastan para proporcionar ningún género de conclusiones sustantivas para la moralidad. Pero supongamos que imponemos una restricción procedimental a esos principios puramente formales. Al igual que las restricciones del espacio y del tiempo puros, que según Kant "no contienen más que simples relaciones" (B66), las restricciones del juego del regateo, incluida la disposición a establecer una aceptación de principios, no contienen detalles específicos de los deseos, intenciones, motivos o creen-

cia de los jugadores³⁰. Rawls dice, sin embargo, que demostrará que, de esas premisas aparentemente estériles, pueden extraerse sólidas y sustantivas conclusiones. En efecto, nos dice que espera demostrar *como un teorema* la proposición de que los agentes racionales, forzados en sus deliberaciones por las normas procedimentales del juego del regateo, así como por los principios puramente formales de la razón práctica, tienen que elegir, necesariamente, sus dos principios de la justicia como la solución de su problema de elección racional colectiva. Es una idea sólida, y, aunque ahora está claro que no surtió efecto, podemos ver fácilmente por qué Rawls persistió en ella hasta el final, a pesar de los muchos reajustes, revisiones y concesiones que transformaron su teoría original.

La segunda semejanza entre Rawls y Kant es mucho menos importante, aunque es, de todos modos, sugestiva. Kant era un inveterado constructor de sistemas. Daba gran importancia a lo que él llamaba la estructura "arquitectónica" de su teoría (ver capítulo III de la *Doctrina Transcendental del Método*, al final de la *Crítica*), y aseguraba siempre que la unidad y la integridad perfectas de su sistema constituían una de las principales pruebas de su verdad. Rawls tiene también una pasión por la totalidad sistemática. Esta pasión se manifiesta en las repetidas referencias hacia delante y hacia atrás, desde una parte de su libro a otra, y en las frecuentes advertencias de que no se intente juzgar la propiedad de ningún elemento particular de la teoría de la justicia, sin valorar su relación con todos los demás elementos, y con la unidad de sus interconexiones.

Ahora bien: la realidad es que, a pesar del genio absoluto de Kant para la categorización y la clasificación, la arquitectura carece virtualmente de valor como guía hacia el verdadero corazón filosófico de sus teorías. Frecuentemente, las apelaciones a la clave de bóveda de la Filosofía Crítica nos apartan de sus más importantes descubrimientos. De un modo análogo, en mi opinión, las pretensiones de Rawls de una perfección sistemática no pueden soportar un gran peso filosófico, aun cuando, como Kant, insista, indudablemente, en que él las propone para que sean consideradas con la máxima seriedad. Como *Teoría de la justicia* es un libro largo, con poca o ninguna paja, y como Rawls es, evidentemente, uno de los más conscientes y sofisticados filósofos que hayan escrito nunca sobre filosofía moral y política, es fácil suponer

³⁰ Esto puede contribuir a explicar por qué considero un refugio tan importante para Rawls la introducción, al menos, de una "sutil" teoría del bien en su modelo.

que, sencillamente, no hemos acertado con el verdadero sentido, cuando alguna parte de su teoría parece no sostenerse en pie. Uno de mis propósitos al exponer la teoría como un desarrollo en varias etapas, más que como un conjunto completo, es el de demostrar mi convicción de que una gran parte de la elaboración arquitectónica es, o bien una especie de sistémica reflexión tardía, o bien, en otro caso, una respuesta defensiva a posibles objeciones formuladas a etapas previas.

XI

EL ANTECEDENTE KANTIANO

EL EXPLÍCITO intento de Rawls de conectar su teoría moral con la filosofía de Kant es la interpretación "kantiana" que ofrece de la posición original en su forma madura, o del velo de la ignorancia. Hablando sucintamente, y un tanto equívocamente, se supone que el velo de la ignorancia tiene como efecto el de limitar a las partes en la posición original a las deliberaciones racionales a que ellas se entregarían como agentes nouménicos más que como criaturas fenoménicas. La verdadera relación entre Kant y Rawls es un poco más complicada, y exigirá algunos antecedentes para aclararla.

La aspiración de Kant en la Filosofía Crítica es la de identificar y aislar el elemento *a priori* en el conocimiento, y la de someterlo a una crítica que establezca su validez. En los párrafos iniciales de la *Fundamentación*, distribuye todo el conocimiento, con una cierta simplificación, entre la Lógica, la Física y la Ética — es decir, entre las normas puramente formales de todo pensamiento en general, los principios del uso teórico de la razón, y los principios del uso práctico de la razón. Las verdades de la lógica alcanzan una validez absolutamente incondicional, universal, en virtud de su total vaciedad. Se aplican, por así decirlo, a todo y a nada. El propósito de la *Crítica de la razón pura* es el de establecer la validez *a priori* de los principios de la razón teórica — física y matemática —, pero, naturalmente, Kant paga un precio por esta victoria. Como la física y la matemática son sustantivas, más que puramente formales, su validez *a priori* es condicionada, más que incondicionada. Puede saberse *a priori* que son verdaderas, a condición de que su campo de aplicación se encierre en los límites de una posible experiencia en general. La sensibilidad, en cuanto facultad del espíritu mediante la cual los objetos de conocimiento pueden ser presentados a la conciencia, fija los límites del conocimiento teórico posible.

Cuando pasa a los principios *a priori* de la razón práctica — a la ética —, Kant abandona deliberadamente la línea de razonamiento que con tanto éxito ha empleado en su tratamiento de la razón teórica. Como los "objetos" de la razón práctica son los fines o metas a que aspi-

ramos en nuestra acción, Kant pudo haber desarrollado una teoría de los principios condicionalmente *a priori* de la razón práctica, según la cual la validez *a priori* de los principios está condicionada por su limitación a los posibles fines o propósitos de los agentes humanos. Una teoría del deseo, análoga a la teoría de las formas de espacio y tiempo, habría proporcionado, entonces, un sistema de principios de razón práctica.

Kant pudo haber tomado esa línea. Si lo hubiera hecho, probablemente habría desarrollado una teoría muy similar a la de Aristóteles o a la de Hume. Pero estaba convencido de que tal teoría, por compleja y sólida que pudiera haber sido, no sería una teoría ética. En el mejor de los casos, sería una teoría de la prudencia racional. Facilitaría conclusiones cuya verdad estaría condicionada por la peculiar estructura del deseo de los seres humanos. De igual modo que nuestra matemática pura sería diferente —sostenía Kant—, si nuestra forma de intuición exterior hubiera sido distinta de lo que es; y de igual modo que nuestra física pura sería diferente, si el sistema de conceptos puros del conocimiento, o categorías, hubiera sido distinto de lo que es; así también tal teoría ética implicaría que nuestros derechos y deberes serían diferentes, si fuese diferente la estructura *de facto* de nuestra facultad de deseo.

La cuestión no consiste en que, para criaturas diferentes de nosotros, la aplicación concreta de principios morales desembocase en diferentes constelaciones particulares de obligaciones. Kant estaba muy satisfecho de garantizar esa inocua pretensión. El problema para él consistía en que, dentro de una teoría ética construida según las líneas de su teoría de conocimiento teórico, los principios morales mismos dependerían de la particular naturaleza de los seres humanos. Hacer tal afirmación —pensaba Kant— equivaldría, simplemente, a revelar que no se comprendía qué son los principios morales.

El problema central de la teoría ética de Kant, por lo tanto, es el de establecer que los principios fundamentales de la moralidad tienen, al mismo tiempo, la dimensión absoluta e incondicionalmente universal de los vacíos principios de la lógica, y la importancia sustantiva, nada trivial, de los primeros principios de la matemática y de la física. Gran parte de la oscuridad y de la confusión de la teoría ética de Kant procede, no de su empleo de un vocabulario psicológico anticuado, ni de las limitaciones culturales de sus convicciones morales pietistas de la Alemania septentrional del siglo XVIII, sino, sencillamente, de la manifiesta dificultad, por no decir imposibilidad, de su meta filosófica.

Kant descubre la clave de su dilema, tanto por lo que se refiere a la ética como al resto de su filosofía, en la distinción entre apariencia y realidad. Cree que, mediante esa distinción, puede resolver el conflicto entre la libertad, que es la marca definidora de la acción moral, y el determinismo, que ha sido establecido para el mundo natural mediante los argumentos de la *Primera Crítica*. Como ser puramente racional, intemporal, como ser-en-sí, yo puedo moverme a actuar en virtud de la comprensión racional de las leyes intemporales de la razón práctica; en resumen, soy libre. Como apariencia en el campo de mi conciencia, como objeto de mi propia razón teórica, estoy subordinado al mismo sistema de leyes empíricas que ordena todos los fenómenos del reino de la apariencia.

Ahora bien: incluso para actuar prudencialmente en la persecución de los fines u objetivos que me ha fijado mi sistema de deseos fenoménicamente determinado, tengo que entregarme a un razonamiento práctico de un género que no es posible para seres simplemente fenoménicos¹¹. Pero los principios de la razón prudencial son condicionales en su forma. Afirman: si tienes como objetivo *E*, haz *A*, que es un medio para *E*. Un agente que realmente tenga el fin *E* como su fin hará *A*, en la medida en que sea racional. Pero, en el análisis de la racionalidad prudencial, nada establece que los agentes, en cuanto racionales, deban adoptar ciertos fines determinados como sus propios fines. Así, Kant puede decir, un tanto equívocamente, pero con el propósito de enlazar su análisis de la razón prudencial con su análisis de la lógica pura, que el principio, "el que quiere el fin quiere los medios", es *analítico*.

Pero los principios morales rigen categóricamente, no hipotéticamente, y el problema de Kant consiste en descubrir alguna forma de establecer la validez objetiva de principios no-vacíos, sustantivos, cuyo campo de aplicación no está, en modo alguno, condicionado por el conocimiento que un agente tenga del fin propuesto por el principio. Kant expresa este problema, un tanto equívocamente, preguntando cómo son posibles *a priori* los principios sintéticos de la razón práctica pura.

En realidad, hay dos problemas distintos que se entrelazan en la

¹¹ A lo largo de esta discusión, recurro muy frecuentemente al detallado análisis de la teoría de Kant en mi comentario sobre la *Fundamentación*. Como algunas de las afirmaciones que yo hago son muy polémicas entre los estudiosos de Kant, yo sólo puedo invitar al lector que las encuentre desconcertantes, o que desee una mayor elaboración de ellas, a consultar *The Autonomy of Reason*.

discusión de Kant acerca de la posibilidad de un imperativo categórico, y si hemos de comprender exactamente cómo la teoría de Rawls se asemeja y no se asemeja a la de Kant, debemos separarlos muy claramente. El primer problema, que domina la discusión de los imperativos en el capítulo II de la *Fundamentación*, es el de cómo una criatura condicionada, finita, como el hombre, puede estar sometida, limitada, por un principio moral incondicionado. Kant aborda esta cuestión como la tarea de investigar "la posibilidad de un imperativo categórico", una investigación de la que él dice que debe realizarse "enteramente *a priori*"³². El conflicto radica en cómo una criatura cuya conducta está determinada por la causalidad natural del deseo y la inclinación puede, sin embargo, estar obligada a determinar sus acciones mediante su comprensión racional de los principios de la razón práctica pura. Estos principios se nos presentan a los hombres como imperativos, porque carecemos de la perfecta o inmaculada voluntad de los seres puramente racionales.

El segundo problema, que motiva mucha de la más sugestiva y difícil argumentación filosófica de Kant, que incluye la derivación del imperativo categórico en los capítulos I y II y la discusión de la humanidad como fin en sí misma, es el de cómo los principios puramente formales de la razón práctica, análogos en su vacía formalidad a las leyes de la lógica, pueden proporcionar principios morales sustantivos. Aquí, la dificultad no reside en cómo unas criaturas finitas y moralmente débiles como nosotros pueden esperar alguna vez adherirse a los principios de moralidad, o en cómo se puede confiar en que lo hagamos. El problema es el de cómo podría haber principios morales sustantivos para que nosotros nos adhiriésemos a ellos, aun suponiendo que fuésemos perfectamente racionales, y, por consiguiente, libres de las tentaciones de la sensibilidad. Para formular la pregunta en los términos que Kant utiliza en la introducción de la *Fundamentación*, ¿de dónde proceden los fines objetivos que pueden servir de contenido a la pura forma de la razón práctica?

Kant ofrece tres respuestas, todas ellas, en última instancia, insatisfactorias. La primera consiste en pretender que, de un principio puramente formal (el imperativo categórico), pueden derivarse precisamente los principios particulares que son universalmente obligatorios para todos los agentes racionales en cuanto tales. La famosa aplicación del imperativo categórico a cuatro "máximas" propuestas (relativas al sui-

cidio, a la falsa promesa, a la ayuda a los otros, y al desarrollo de los propios talentos naturales) es el esfuerzo —fracasado— de Kant por demostrar que el imperativo categórico puede excluir las máximas de acción erróneas y regir en las verdaderas. La segunda tentativa —un poco más adelante, en el mismo capítulo— es la doctrina de que la humanidad es un fin-en-sí-misma, y, por lo tanto, un fin que todo agente racional debe tomar como su fin. Aunque la discusión de la dignidad de la humanidad es uno de los más vivos pasajes de la literatura de la filosofía moral, su argumentación no puede establecer la conclusión que Kant necesita, y que es la de que hay alguna meta o algún fin que todas las criaturas racionales en cuanto tales tienen como su fin, y que, por consiguiente, da contenido a la vacía fórmula del imperativo categórico. La tercera y última tentativa de Kant, en la *Metafísica de las costumbres*, consiste, sencillamente, en afirmar sin argumentación que hay dos objetivos o fines "obligatorios": concretamente, la felicidad de los otros y mi propia perfección.

Por último, hemos de concluir que Kant fracasó a la hora de descubrir un modo de deducir fines objetivos, obligatorios, del simple análisis de lo que tiene que ser un agente racional. En consecuencia, no pudo establecer tampoco la validez incondicionalmente universal de ningún principio sustantivo de razón práctica. En mi opinión, su fracaso era inevitable, porque no hay tales principios, pero, en todo caso, estoy convencido de que los argumentos de Kant no son válidos.

³² *Kants Werke*, Ak., IV, 419-20.

problema es el de cómo unas criaturas con intereses propios como nosotros pueden adoptar y adherirse a un principio de razón práctica que evita toda referencia a los objetos de nuestros intereses. El segundo problema es el de cómo un principio puramente formal de razón práctica puede proporcionar conclusiones sustantivas en la forma de principios morales no-vacíos.

Kant dice que ser autónomo es hallarse "sometido solamente a leyes que están hechas por (uno mismo) y que, sin embargo, son universales" (Ak., IV, 432). Una lectura superficial de éste y de otros pasajes parece indicar que la esencia de la autonomía es la *auto-legislación*. Pero, mirado más detenidamente, resulta claro que Kant piensa en la distinción entre legislación o voluntad guiada por el deseo o la inclinación, y, en consecuencia, sometida a éstos —lo que él denomina como heteronomía—, y legislación o voluntad que ignora las motivaciones sensoriales o que hace abstracción de ellas. Kant llama autónomo a ese tipo de legislación. La conexión entre autonomía y universalidad consiste, sencillamente, en que cuando yo quiero un principio, puramente en mi condición de agente racional, con abstracción de todo contenido sensorial particularizador, quiero, necesariamente, la misma ley que sería legislada por cualquier otro agente racional en iguales circunstancias. Como es el contenido de la voluntad o del deseo lo que diferencia a un agente racional de otro, y no la forma pura de la razón práctica, se sigue que la ley que yo quiero en cuanto agente racional es la ley que todo agente racional en cuanto tal querría, y, en consecuencia, es obligatoria para todos los agentes racionales.

Así, el utilitarista que se da *a sí mismo* la ley: "Actúa de modo que produzcas la mayor felicidad al mayor número" es heterónomo, más que autónomo, porque voluntariamente toma como fin de su acción el placer o la felicidad, ya sean los suyos o los de otro. La ley que él quiere no es, por lo tanto, obligatoria para un agente racional que no se proponga ese fin. En el mejor de los casos, ese principio, aunque altruista, sería un principio condicionado, más que incondicionado, y su expresión adoptaría la forma de un imperativo hipotético, más que la de un imperativo categórico.

Toda vez que es en cuanto agentes nouménicos como somos capaces de voluntad racional, y es de nuestro carácter fenoménico del que pueden predicarse las particularidades de nuestro carácter espacio-temporal, nos abstraemos, en nuestra función de agentes racionales, de todo lo que diferencia a uno de nosotros de otro, en el campo de la apariencia. Este es el hecho que conduce a Rawls a asociar su descripción

de la posición original con el reino de los fines, y a decir que una parte en la posición original es como un agente nouménico.

Pero, cuando las partes en la posición original son privadas de todo conocimiento de sí mismas como agentes particulares, son privadas también de la base para una deliberación racional, *exactamente como ocurre con los agentes nouménicos de Kant*. Mientras un agente racional atiende sólo a la forma pura de la razón práctica, acaso no pueda llegar a conclusiones sustantivas respecto a los principios de una acción justa. Las partes que se encuentren bajo un total velo de la ignorancia, adolecerán exactamente de la misma incapacidad. Así, Rawls les permite saber, primero, que son racionalmente *auto-interesados*, y no sólo que son racionales. Como tampoco esto es suficiente para que puedan llegar a conclusiones importantes, Rawls agrega un conocimiento de los hechos básicos de la sociedad humana, un conocimiento de que ellos poseen planes de vida de una cierta y determinada estructura, y una teoría de bienes primarios como suplemento de la teoría de los planes de vida. Rawls cree que, al fin, las partes cuentan con lo suficiente para extraer algunas conclusiones.

En la cuarta parte, discutiré que Rawls ha dado a sus partes demasiado (el conocimiento de los hechos básicos...) y demasiado poco (la simple noción de un plan de vida y una inadecuada teoría de bienes primarios) para el problema de elección con que se enfrentan. Pero, aun cuando su acopio de conocimiento fuese, como diría Goldilocks, suficientemente correcto, tampoco proporcionaría principios "autónomamente queridos", en el sentido kantiano. Lo que Rawls pretende es que "el velo de la ignorancia priva a las personas en la posición original del conocimiento que les capacitaría para elegir principios heterónomos" (288), pero, en realidad, eso solamente garantiza que sus principios serán, por así decirlo, generalmente heterónomos, más que particularmente heterónomos. La elección de principios está motivada por el propio interés, más que por la Idea del Bien.

El verdadero mérito de la interpretación kantiana de Rawls radica en su elaboración del segundo problema enlazado con el imperativo categórico —el problema de explicar cómo una criatura sensorialmente limitada puede estar bajo unos principios categóricos de moralidad, y obligada por ellos. Kant, realmente, no tiene una respuesta, a pesar de su elaborada distinción metafísica entre los planos de la apariencia y de la realidad, y tampoco de Rawls puede decirse que tenga una respuesta, en sentido estricto. De todos modos, la noción de Rawls de la "expresión" (291), y con ella su pretensión de que, adoptando el pun-

to de vista de la posición original, "expresamos nuestra naturaleza como seres libres y racionales" (293), articula, en mi opinión, una de las ideas centrales de la filosofía moral de Kant.

Lo cierto es que Rawls, al igual que Kant, no puede ofrecer réplica alguna al escéptico que pregunte: "¿Por qué tengo que ser moral?". Pero si la argumentación a partir de la posición original fuese correcta en otros aspectos, ese hecho sería, en el mejor de los casos, una débil objeción.

Por último, podemos comprender la íntima conexión entre la prioridad de la libertad y la interpretación kantiana de la posición original. Las formas políticas demandadas por el principio de libertad igual son la encarnación institucional de la decisión colectiva de dar expresión al recíproco reconocimiento de los hombres como personas racionales libres e iguales. Ese es el principio constitucional fundamental del Reino de los Fines. Rawls se equivoca (a mi juicio), al intentar cargar ese principio con planes de vida y con una sutil teoría del bien. Creo que también se equivoca, al intentar extraer el mutuo reconocimiento de la dignidad de la personalidad, de una presión del regateo sobre un auto-interés racional. Pero, dado el fracaso de Kant al no alcanzar unas conclusiones sustantivas satisfactorias sobre la base de un análisis de la forma pura de la razón práctica, los atractivos de la táctica de Rawls no son difíciles de comprender.

CUARTA PARTE CRÍTICA DE LA TEORÍA

En esta parte de la obra, el autor se propone la siguiente tarea: la de explicar la posición original de un punto de vista kantiano, y mostrar que esta posición original es la única que puede ser justificada por la razón práctica. El autor comienza por explicar la posición original de un punto de vista kantiano, y luego se propone la tarea de mostrar que esta posición original es la única que puede ser justificada por la razón práctica. El autor comienza por explicar la posición original de un punto de vista kantiano, y luego se propone la tarea de mostrar que esta posición original es la única que puede ser justificada por la razón práctica.

En esta parte de la obra, el autor se propone la siguiente tarea: la de explicar la posición original de un punto de vista kantiano, y mostrar que esta posición original es la única que puede ser justificada por la razón práctica. El autor comienza por explicar la posición original de un punto de vista kantiano, y luego se propone la tarea de mostrar que esta posición original es la única que puede ser justificada por la razón práctica. El autor comienza por explicar la posición original de un punto de vista kantiano, y luego se propone la tarea de mostrar que esta posición original es la única que puede ser justificada por la razón práctica.

En esta parte de la obra, el autor se propone la siguiente tarea: la de explicar la posición original de un punto de vista kantiano, y mostrar que esta posición original es la única que puede ser justificada por la razón práctica. El autor comienza por explicar la posición original de un punto de vista kantiano, y luego se propone la tarea de mostrar que esta posición original es la única que puede ser justificada por la razón práctica.

se despierta de la posición original. "Representamos entonces a los individuos como seres racionales y libres" (1971, p. 13). En esta sección, Rawls expone su teoría de la justicia, la cual se basa en la posición original.

Los principios de la justicia, al igual que la teoría de la justicia, se basan en la posición original. En esta sección, Rawls expone su teoría de la justicia, la cual se basa en la posición original.

En esta sección, Rawls expone su teoría de la justicia, la cual se basa en la posición original. En esta sección, Rawls expone su teoría de la justicia, la cual se basa en la posición original. En esta sección, Rawls expone su teoría de la justicia, la cual se basa en la posición original.

XIII

LOS HECHOS GENERALES ACERCA DE LA SOCIEDAD HUMANA

EN LA SEGUNDA PARTE, al intentar la exposición de la teoría de Rawls como el resultado de un proceso de desarrollo, he esbozado un cierto número de críticas de las primeras formas del modelo y sugerido que podríamos comprender mejor la teoría final como una respuesta, en parte, a las dificultades reveladas por esas críticas. La mayoría de las dificultades, si no todas, fueron vistas por el propio Rawls; su desarrollo no fue tanto una respuesta a los críticos como un crecimiento intelectual o filosófico interior. Al final de la segunda parte, en mi discusión de la prioridad de la libertad y del principio de los ahorros, he indicado algunas razones para cuestionar esos elementos de la teoría final, pero, como son elaboraciones secundarias, no creo que la filosofía de Rawls, en modo alguno, se mantenga o se derrumbe a causa de su justificabilidad.

Ahora es el momento de considerar la teoría como tal, y de preguntarse si puede ser adecuadamente defendida. En las cuatro secciones de esta parte del ensayo, ofreceré un cierto número de argumentos destinados a demostrar que la teoría de Rawls es incorrecta. En cuanto a la tesis central del ensayo, me dedicaré principalmente a lo que considero que es el núcleo de la teoría, concretamente el modelo de un juego del regateo jugado en las condiciones que Rawls resume como "la posición original". En esta sección y en la siguiente, discreparé de algunos de los reajustes *ad hoc* de la posición original, que Rawls introdujo para mantener en funcionamiento su juego del regateo. En especial, discutiré que las condiciones de conocimiento de la posición original son imposibles, que la "sutil teoría" de los bienes primarios es inadecuada a las necesidades de Rawls, y que la noción de un plan de vida es una distorsión de lo humanamente racional.

En el capítulo XV, intentaré precisar la estructura del juego del regateo, a fin de determinar si puede esperarse, razonablemente, desembocar en la "solución" propuesta por Rawls. Como veremos, la respuesta es negativa, a causa de una variedad de lo que me parece que son razones muy serias. Rawls es confuso acerca de lo que quiere decir

con "maximin"; la regla maximin de elección de estrategias no proporciona por sí sola los dos principios de la justicia en ninguna reconstrucción aceptable del juego; la invocación de Rawls de lo que él llama "justicia procesal pura" es incorrecta, de un modo que es importante y perjudicial para su teoría; y sus argumentos en favor de la adopción de la regla maximin, sencillamente, no se sostienen en pie.

En la sección final de esta parte, trataré de abordar la intrincada cuestión del *status* lógico de la argumentación de Rawls, con especial atención a su característica doctrina del "equilibrio reflexivo". El mensaje de la cuarta parte es el de que el núcleo de la teoría de Rawls es erróneo: sencillamente, los dos principios de la justicia no son, en ningún sentido, la solución a la clase de juego del regateo esbozado por Rawls.

Veamos, en primer lugar, las condiciones del conocimiento en la posición original³⁴. Bajo el velo de la ignorancia, una parte olvida todos los detalles de su vida personal, incluidos sus talentos, facultades, intereses, objetivos, cultura, sexo, edad, plan de vida, situación histórica, etc. Por lo que se refiere a los detalles, sólo sabe que, fuera del templo de la justicia, espera una situación limitada o definida por las circunstancias de la justicia. Pero conserva su conocimiento de los "hechos generales acerca de la sociedad humana". Comprende "las cuestiones políticas y los principios de la teoría económica; conoce(n) las bases de la organización social y las leyes de la psicología humana" (164). ¿Es esto, en principio, epistemológicamente posible?

Prescindamos inmediatamente de algunas objeciones más o menos disparatadas. El velo de la ignorancia es un recurso literario destinado a dar brigen a una pretensión lógica. La pretensión es, sencillamente, la de que, en nuestro razonamiento acerca de las cuestiones morales y sociales, podemos optar por poner en práctica la misma abstracción de las particularidades que hemos aprendido a utilizar en nuestro razonamiento matemático, por ejemplo. Así como Rawls no nos pide que creamos que haya habido nunca una persona totalmente exenta de envidia, tampoco nos exige que imaginemos a una criatura poseída solamente de conocimiento general.

³⁴ Entre los muchos artículos que tratan de este tema, permítaseme dirigir la atención del lector, especialmente, hacia el ensayo excepcionalmente complejo y penetrante de Benjamin Barber, "Justifying Justice: Problems of Psychology, Politics and Measurement in Rawls", en Norman Daniels, *Reading Rawls* (Basic Books, 1975). Su ensayo trata también los temas discutidos en el capítulo XV.

Tampoco la cuestión genética de las fuentes y del desarrollo de nuestro conocimiento es directamente relevante — al menos, no lo es en ningún sentido trivial. El conocimiento general ha sido adquirido, indudablemente, por inducción de los detalles, y por reflexión sobre los mismos, incluidos los detalles de la propia experiencia de la parte; Rawls no niega, en modo alguno, esta verdad evidente.

Pero las partes en la posición original *son* racionales, y se supone, por lo tanto, que conocen todo lo que puede deducirse, o, en otro caso, inferirse razonablemente, de su acopio básico de conocimiento. No pueden, por así decirlo, saber que todos los hombres son animales y que todos los animales son mortales, y olvidar, sin embargo, que todos los hombres son mortales.

Ahora podría objetarse que el estudio del hombre todavía tiene que proporcionar *leyes* del tipo de las que ofrecen las ciencias naturales. Un conocimiento general como el que poseemos de las bases de la organización social y de las obras de la psicología humana es una pobre fruslería, si se compara con las fórmulas matemáticas de la física, de la química, o incluso de la biología. Pero, aunque hay, indudablemente, mucho de verdad en este punto, no puede constituir una objeción a la teoría de Rawls, ni en su versión original, ni en la versión final del velo de la ignorancia. Se supone que los participantes en el juego del regateo son racionales, no omniscientes. Saben todo lo que puede imaginarse que unas personas inteligentes e ilustradas saben acerca del hombre y de la sociedad. Si es mucho, sus deliberaciones serán más fructíferas; si es muy poco, tendrán que hacer uso de los cálculos que su conocimiento permita. Pero, como en este mundo sólo se puede exigir de un agente racional una rectitud subjetiva, no puede culparse a Rawls, sencillamente, del inadecuado desarrollo de las ciencias sociales.

Sin embargo, si hay serias razones metodológicas o epistemológicas para suponer que los seres humanos *no podrían* tener los tipos de conocimiento general que Rawls atribuye a las partes en la posición original, sin tener también que ser conscientes de los tipos de hechos particulares acerca de sí mismos que el velo de la ignorancia les oculta — en resumen, si la particular combinación de conocimiento e ignorancia requerida por la construcción de Rawls es, en principio, imposible —, entonces habrá que cuestionar la teoría en su totalidad. Ni por un momento supongo que pueda yo demostrar tan terminante pretensión (aunque Rawls no ha dicho nada, en absoluto, que demuestre lo contrario), pero creo que hay poderosas razones para dudar, por lo menos, de la posición original, tal como Rawls la ha caracterizado.

Consideremos, en primer lugar, el conocimiento de los asuntos políticos, los principios de la teoría económica y las bases de la organización social. La teoría de Rawls da por sentado que ese conocimiento es ahistórico en sus fundamentos epistemológicos, que tiene el mismo carácter trans-temporalmente impersonal que poseen las verdades de las ciencias naturales. Esto no quiere decir que el conocimiento de la teoría económica o de la organización social sea estático, que no guarde relación con los procesos de desarrollo o de cambio históricos. Las partes en la posición original, según Rawls, podrían saber, por ejemplo, que la economía de una sociedad tiene que desarrollarse, necesariamente, a través de una determinada serie de etapas, cada una de ellas definida por las relaciones de producción sociales dominantes (si ese *es*, realmente, un hecho básico de la sociedad). O podrían saber —para recurrir a una alusión clásica— que una sociedad justa y bien ordenada es propensa a degenerar, primero, en un estado militar, después en una oligarquía, a continuación en una democracia caótica, y finalmente en una tiranía despótica (cf. *La República* de Platón). Pero, para los propósitos de Rawls, es necesario suponer que ese conocimiento, en la medida en que *es* conocimiento, es intemporal, de donde se sigue que, al menos, sería imaginable que un hombre pudiera poseerlo, en cualquier espacio o en cualquier tiempo en que existiese. Desde luego, todo conocimiento empírico se basa en la previa experiencia propia y ajena, de modo que una parte en la posición original que se encontrase en posesión de tal conocimiento podría inferir que ella no era la primera persona en la historia del universo. Pero, como el conocimiento (continuando con la imagen) podría ser un legado de una civilización desaparecida, esa persona, en opinión de Rawls, no podría inferir que ella no pertenecía a la primera generación de *su* sociedad, ni podría dar por supuesto que su sociedad hubiera experimentado todo lo necesario para adquirir tal conocimiento.

Pero yo quisiera sugerir que la concepción de Rawls de la naturaleza del conocimiento de la sociedad es errónea. Siguiendo a Marx, en la tradición de la disciplina habitualmente denominada como Sociología del Conocimiento, yo señalaría que nuestro conocimiento de la realidad social es fundamentalmente diferente del conocimiento de la realidad natural.

La sociedad, o realidad social, es un producto humano colectivo, de un modo que la naturaleza objetiva no lo es. La sociedad es la suma total de los sentimientos, esperanzas, hábitos, modelos de interacciones, y creencias de los hombres y de las mujeres que la componen. La

existencia y la persistencia de la sociedad requieren la sistemática hipótesis y objetivación de lo subjetivo: la percepción como independientemente reales, por parte de los hombres y de las mujeres, de unas funciones y relaciones sociales que, en realidad, dependen de aquellos que las ocupan. Así, pues, la sociedad, según he señalado en otra parte, es una especie de *folie à tous*. Es una errónea percepción sistemática, colectiva, o una falsa conciencia, que simultáneamente *expresa* el grado de comprensión o de incompreensión de los hombres y de las mujeres de un momento dado de la historia de la sociedad y *configura* también sus sentimientos, su conducta y sus expectativas en relación con el mantenimiento o con la alteración de esa realidad social.

El signo que define la falsa conciencia colectiva es la creencia de que la sociedad es un *objeto* gobernado por leyes inmutables. Se considera que esas leyes pueden ser exploradas, descubiertas, formuladas y puestas al servicio de los propósitos humanos, de igual modo que pueden serlo las leyes de la naturaleza; pero, como leyes de una realidad objetiva, no pueden ser *cambiadas*. Se cree que son como las leyes de la física. Colocados ante una roca que avanza, podemos apartarnos de su ruta, *evitando* así las consecuencias de las leyes del movimiento; o podemos desviar la roca, mediante la colocación cuidadosamente calculada de obstáculos, *utilizando* así las leyes del movimiento; pero no podemos, mediante ninguna suma de meditación, de creación de conciencia o de auto-examen colectivo, *abolir* las leyes del movimiento y anular así la amenaza de la roca. De igual modo, podemos tratar de evitar las leyes de la oferta y la demanda, o utilizar las leyes de la oferta y la demanda, pero no podemos abolir las leyes de la oferta y la demanda. Hay "hechos generales de la sociedad humana", y es un mínimo signo de racionalidad el de reconocer la objetividad de lo real. O así lo presupone el recurso del "velo de la ignorancia" de Rawls.

Pero, en mi opinión, la verdad es que nuestro conocimiento de la sociedad es diferente, por su carácter, de nuestro conocimiento de la naturaleza, porque el objeto de nuestro conocimiento de la sociedad es diferente, por su carácter, del objeto de nuestro conocimiento de la naturaleza. Nuestro conocimiento de la sociedad es un proyecto humano colectivo, un proyecto de auto-comprensión y *también de auto-alteración*. Para que la teoría económica clásica se desarrollase como un cuerpo de conocimiento acerca de las actividades del mercado, fue necesario, no sólo que un determinado número de pensadores extremadamente sutiles llevasen a cabo importantes observaciones y poderosas abstracciones, sino también que los modelos de interacciones socio-

económicas experimentasen cambios fundamentales. Fue necesario que las dimensiones culturales, religiosas y políticas de la actividad "económica" fuesen separadas, no sólo por los observadores, por los teóricos, sino también por los participantes en aquellas interacciones. Tenían que cambiar incluso las formas de la organización de la personalidad. En resumen, la teoría económica clásica no fue un *descubrimiento* de leyes de acción intemporal, a la manera en que lo fue la mecánica newtoniana. Más bien, fue una expresión y una reflexión de una nueva forma de organización social que había sido constituida, creada, por las acciones e interacciones de innumerables individuos. Los economistas clásicos no se dieron cuenta de este hecho, naturalmente. Creían que estaban *descubriendo* las leyes del mercado. La ulterior comprobación de la verdadera condición epistemológica de aquellas "leyes" fue, en sí misma, un proyecto social colectivo.

Para mi "comprensión" de los "descubrimientos" de la economía clásica, tengo que concebir la sociedad como la concibieron los economistas clásicos. De otro modo, no podría suponer que estaban describiendo el "mundo real". Y como la sociedad, al contrario de la naturaleza, no es un objeto de existencia independiente —como, en realidad, cada uno de nosotros desarrolla, al mismo tiempo, una personalidad coherente, mediante la internalización de una concepción social de la realidad social, y luego mantiene (o altera) esa realidad social mediante la expresión de esa concepción social en interacciones sociales—, se sigue que yo creo que una teoría económica es *conocimiento* de la sociedad, solamente en la medida en que expresa la concepción de la realidad social que corresponde a mi etapa de desarrollo, y a la de mi sociedad, en la progresiva superación de la falsa conciencia colectiva.

Un razonamiento análogo podría hacerse en relación con el conocimiento general de la naturaleza. Yo no puedo ser auténticamente neutral entre una concepción científica y una concepción animista del mundo natural. Con esto no quiero decir, simplemente, que las dos concepciones de la naturaleza sean incompatibles, de modo que un hombre racional tendría que reconocer la necesidad de elegir entre ellas. Quiero decir, más bien, que las orientaciones científica y animista respecto a la naturaleza son formas incompatibles de conocer y de ser, de modo que nadie podría entregarse, auténticamente, a una deliberación racional respecto a cuál de las dos debería elegir. (Es el mismo tipo de cuestión que Kierkegaard suscita con tan brillante fuerza satírica en sus ataques a la posición "objetiva" respecto a la fe religiosa.)

Como he distinguido tan terminantemente entre los objetos de la ciencia natural y los "objetos" de la ciencia social, podría suponerse que niego la condición de conocimiento objetivo o científico a las creencias acerca de la sociedad. No, en absoluto. De igual modo que estoy seguro de que la aproximación científica a la naturaleza es superior a la animista, así también tengo la seguridad de que la historia científica es superior a la anécdota, de que la economía clásica fue un avance respecto a la teoría medieval del justo precio, y de que la sociología del conocimiento es, en la medida en que hasta hoy se ha desarrollado, una victoria de la inteligencia humana en la lucha por el conocimiento. Pero, mientras el conocimiento natural es intemporal, el conocimiento social es histórico, auto-reflexivo, y constitutivo a la vez que descriptivo.

Desde luego, estas observaciones son muy generales y especulativas, pero, si tienen algún valor, será el de cuestionar la descripción de Rawls de las condiciones del conocimiento en la posición original. Me parece evidente que las partes en la posición original son hombres y mujeres racionales, temporales, científicos. Este hecho debe ser conocido por ellas, pues así se deduce del modo en que analizan los problemas de elección y en que razonan acerca de las alternativas. Aunque Rawls da mucha importancia a la tolerancia religiosa y dice que es discutible en la posición original si las partes tienen creencias religiosas, y, en caso afirmativo, cuáles son, esto no es más que una concesión a la tradición liberal del pluralismo. Si una parte en la posición original sabe que tiene un plan racional de vida, entonces sabe que ella no es Søren Kierkegaard, ni Moisés, ni san Pablo, aunque sería muy posible que fuese Carlos Marx, o Ramsés, o César Augusto. De igual modo, si los "hechos generales de la sociedad humana" incluyen una conciencia del carácter ideológico de la teoría económica clásica, entonces una parte en la posición original puede inferir que ella tiene que vivir en una sociedad que ha superado, en su avance, las etapas de la caza y de la recolección y de la agricultura primitiva, las primeras etapas de la racionalización de la producción industrial, e incluso las primeras etapas de la formación de una economía capitalista. Sabrá que su sociedad ha alcanzado o superado una etapa en la que hay una separación funcional y una institucionalización de los procesos de adquisición y desarrollo del conocimiento, juntamente con un conflicto entre las clases, en el que la clase dominante cuenta con los servicios de unos intelectuales para producir una racionalización de lo que es una injusta distribución de los ingresos y de la riqueza, pero que todavía no es consi-

derado así por todos los miembros de la sociedad. Sabrá que *su propia* sociedad ha alcanzado esta etapa, y no sólo que *una* sociedad humana, en el pasado, ha adquirido tal conocimiento, porque las pretensiones de conocimiento formuladas por el análisis ideológico de la teoría económica no se le presentarán como pretensiones aceptables o comprensibles acerca de la sociedad humana, a no ser que su propia sociedad haya progresado hasta una cierta etapa en la progresiva desmixtificación de las relaciones sociales.

Los párrafos anteriores son, evidentemente, afirmaciones, más que pruebas. Pero no son, en modo alguno, afirmaciones temperamentales ni nuevas. Representan, más o menos exactamente, una tradición dominante en la teoría social del pasado siglo y medio. Los fáciles supuestos de Rawls acerca de las condiciones del conocimiento en la posición original ignoran, simplemente, los argumentos y los análisis de esta tradición, en favor de otra más antigua, que yo considero menos refinada y menos defendible. En todo caso, creo que no pueden aceptarse el velo de la ignorancia y la posición original como construcciones analíticas aceptables, a efectos de explicar y defender una teoría de la justicia, mientras no se haya resuelto esta disputa epistemológica relativa a la naturaleza de nuestro conocimiento de la sociedad.

Un problema análogo es el planteado por el concepto de una sociedad bien ordenada, que ya hemos considerado brevemente. El signo distintivo de una sociedad bien ordenada es la *publicidad* —“cada cual acepta y sabe que los otros aceptan los mismos principios de justicia” (21). Rawls afirma que él trata de limitar su discusión a “los principios de la justicia que regularían una sociedad bien ordenada” (25). Describe esto como una decisión de tratar “una teoría de la obediencia total como opuesta a la de la obediencia parcial”, lo que parece implicar que la diferencia entre una sociedad bien ordenada y una sociedad no bien ordenada radica en el grado en que los individuos conocen, observan y saben que otros observan los principios de la justicia. En capítulos siguientes, Rawls trata como una cuestión discutible si una sociedad gobernada de un modo bien ordenado por sus dos principios de la justicia tendrá una economía socialista o una economía capitalista. Pero supongamos que Marx tiene razón en cuanto a la relación entre la sistemática explotación de clase y la existencia de la falsa conciencia y la ideología en una sociedad (como yo supongo, en algún sentido, que la tiene). Supongamos que la propiedad privada y el control de los medios de producción, y la producción por la ganancia más que para las necesidades, son socialmente intolerables sin amplias y sis-

temáticas concepciones erróneas acerca de la verdadera naturaleza de las relaciones económicas, de poder y políticas. En consecuencia, sería imposible, en principio, que una sociedad bien ordenada tuviese una estructura económica capitalista, porque, en el momento en que se hubieran alcanzado las condiciones de publicidad requeridas por el concepto del buen ordenamiento —en realidad, antes de que hubieran sido alcanzadas plenamente—, habría tenido lugar un cambio revolucionario en las relaciones sociales de producción.

Una vez más, me apresuro a reconocer que simplemente he sugerido una posibilidad, más que discutido —y, mucho menos, demostrado— una hipótesis. Pero incluso esta sugerencia, respaldada como lo está por un siglo de importante teoría social, basta para demostrar que Rawls ha acumulado supuestos extremadamente sólidos —y me atrevo a decir que ideológicamente influidos— en su caracterización de la posición original bajo el velo de la ignorancia.

Cuando atendemos a la afirmación de Rawls de que las partes en la posición original conocen las “leyes de la psicología humana”, encontramos análogos problemas metodológicos y epistemológicos³³. También aquí, la dificultad radica en el fácil supuesto de Rawls de que el conocimiento de la personalidad humana, como el conocimiento de los cuerpos en movimiento, es intemporal e impersonal. Ya sé que algún supuesto similar es ampliamente compartido por muchas escuelas de psicología social y de teoría de la personalidad, pero quiero hacer constar que es muy señaladamente *no* compartido por Freud y por los que le han seguido en la tradición psicoanalítica. Al apelar a los métodos y enfoques de Freud, me situó, una vez más, en la posición de expresar, simplemente, mis compromisos y preferencias intelectuales. Debo rogar al lector que considere estas observaciones como lo que son, y que consulte a sus propias convicciones en estas materias.

La clave del descubrimiento de Freud es, en mi opinión, su descubrimiento del inconsciente. Simplificando mucho, podemos decir que nuestras mentes están trabajando siempre en dos planos, el consciente y el inconsciente. Cada uno de nosotros capta y responde a sugerencias o mensajes de los otros, en ambos planos; y cada uno de nosotros se comunica con los otros, consciente e inconscientemente. A través de los procesos dinámicos de la proyección, de la transferencia, del desplazamiento y de la represión, y a través del despliegue de defensas

³³ Las páginas siguientes deben toda la coherencia y la lógica que puedan tener a la ayuda de mi mujer, la profesora Cynthia Griffin Wolff.

elaboradamente auto-engañosas, transformamos, ocultamos o negamos deseos, pensamientos y sentimientos penosos, o vergonzosos, o peligrosos. Ninguna descripción de la experiencia y de la conducta humanas podría ser satisfactoria, sin el reconocimiento de esta rica complejidad de lo consciente y de lo inconsciente.

Ahora bien: al tratar a un paciente, el psicoanalista le responde emocionalmente —le responde como ser humano—, además de observarle clínicamente. El psicoanalista utiliza sus respuestas emocionales espontáneas como guías para las comunicaciones latentes que proceden del inconsciente del paciente. Si el psicoanalista niega o reprime sus propias respuestas emocionales en beneficio de la objetividad científica o de la impersonalidad clínica, se priva de un indispensable instrumento de investigación y de interpretación.

Pero más importante aún, según psicoanalistas posteriores han comprobado, es que, en la situación terapéutica, tiene lugar un proceso de "contra-transferencia". En una terapia afortunada, no sólo el paciente transfiere sus fantasías al analista, sino que también el analista experimenta una transferencia recíproca de sus propias fantasías (que él, como ser humano, debe albergar también en su inconsciente). Tampoco aquí la vía de la observación y del conocimiento avanza a través de la supresión, sino más bien de la aceptación auto-reflexiva por parte del analista de esos procesos en sí mismo.

Hasta ahora, cabría suponer que Rawls podría aceptar fácilmente todo lo que yo he dicho, suponiendo que él esté tan influido como yo por los enfoques y por los métodos del psicoanálisis. Porque es sólo en cuanto participantes en la posición original como las partes ignoran sus personalidades particulares y sus biografías. Es perfectamente coherente con la teoría de Rawls afirmar que su adquisición de un conocimiento de las "leyes de la psicología humana" las ha implicado, personal y directamente, en interacciones auto-reflexivas con otros seres humanos. Pero razonar de este modo equivaldría a dar por sentado que, una vez que un individuo ha llegado a un conocimiento de una determinada persona, o de la patología emocional en general, puede poner entre paréntesis su auto-conocimiento para dejar en pie un cuerpo de conocimiento científico objetivo, impersonal. Considero, por el contrario, que el conocimiento alcanzado a través de un reconocimiento, una aceptación y una trascendencia de las propias proyecciones y transferencias conserva la marca de sus orígenes. La interpretación de un analista joven de las fantasías, defensas y emociones de un paciente viejo diferirá, necesariamente, de la interpretación —muy posiblemente

te, de igual fortuna desde el punto de vista terapéutico— de un analista mayor. Un analista masculino tendrá una interpretación de la significación del embarazo y del parto, diferente de la que tendrá una analista femenina. Y, aunque la edad y el sexo no tienen el mismo carácter que el color como diferenciadores de los seres humanos, un analista negro, en nuestra sociedad, tendrá una interpretación de la significación emocional de la raza, diferente de la que tendrá un analista blanco.

Yo no digo que solamente los viejos puedan comprender a los viejos, solamente los hombres puedan comprender a los hombres, solamente las mujeres a las mujeres, solamente los blancos a los blancos, y solamente los negros a los negros. Lo que digo es que cada uno de nosotros comprende a los demás, ya sea informalmente en nuestros contactos cotidianos, ya sea clínicamente como expertos terapeutas, a través de la mediación de nuestra propia composición emocional fundamental. En consecuencia, nuestro conocimiento ostenta la marca del conocedor, así como la marca del conocido. Asegurar que las partes en la posición original conocerán las leyes de la psicología humana sin conocer quiénes son, y si son viejos o jóvenes, varones o hembras, blancos o negros, homosexuales o heterosexuales, equivale a decir que *comprenderán* realmente las leyes de la psicología humana, y no simplemente que emitirán unas frases vacías. Y eso, en mi opinión, no es posible, si Freud y la tradición psicoanalítica están en lo cierto.

Estas observaciones acerca de los "hechos generales de la sociedad humana" constituyen una amplia queja contra el menosprecio de Marx y de Freud por parte de Rawls. Se han escrito bibliotecas enteras acerca de cada una de estas figuras y de las tradiciones intelectuales que ellas iniciaron. Yo estoy, desde luego, intelectualmente comprometido con esas tradiciones, y espero que los lectores que compartan mis inclinaciones estarán de acuerdo en que Rawls no les ha hecho justicia. Pero, para aquellos otros lectores que duden o rechacen las orientaciones de Marx y de Freud, yo sugeriría que algo hay de sospechoso en una construcción teórica, destinada a captar los caracteres esenciales de la racionalidad deliberativa, si excluye *ab initio* a Freud y a Marx.

XIV

BIENES PRIMARIOS Y PLANES DE VIDA³⁶

ADEMÁS DE su conocimiento de los hechos generales de la sociedad humana y del hecho particular de que su sociedad presenta las circunstancias de la justicia, se dice que las partes en la posición original saben que, como seres humanos racionales, tienen planes de vida coherentes, integrados. Asimismo, aunque no conocen los detalles de esos planes, saben que los planes requieren ciertas clases de bienes para su satisfactoria realización. Esos "bienes primarios" son de tal clase que cada parte, a pesar de ignorar los detalles de su plan de vida, sabe que prefiere tener más que menos. Como hemos visto, estos supuestos, por abstractos que puedan parecer, tienen un gran peso sustantivo —o, quizá más exactamente, ejercen presiones sustantivas sumamente estrechas sobre el problema de elección con que se enfrentan las partes en la posición original.

En la primera formulación de la teoría de los bienes primarios, Rawls los identifica como pertenecientes a las categorías generales de "derechos, libertades y oportunidades, así como ingresos y riquezas" (114). Posteriormente, revisa esta descripción para dar prioridad a los derechos y a las libertades, y altera un tanto la redacción y la interpretación de los dos principios para expresar este ordenamiento. Ya he tratado la prioridad de la libertad en la segunda parte. Por consiguiente, fijemos ahora nuestra atención en la noción de un plan de vida y en su exigencia de "oportunidades, así como ingresos y riquezas".

Ante todo, propongo dejar a un lado cualquier tipo de objeciones a la teoría de Rawls basadas en la posibilidad de planes de vida ascéticos o cuasi-religiosos. Naturalmente, si un hombre ha tomado como lema el mandamiento de Thoreau —"Simplificar, simplificar"—, entonces, más allá de un modestísimo nivel, no tendrá una utilidad marginal positiva para los bienes primarios (con excepción de los derechos

³⁶ Con esta sección, ver especialmente Michael Teitelman, "The Limits of Individualism", *J. Phil.*, 69 (1972), y la réplica de Rawls a esa y a otras críticas en "Fairness to Goodness", *Phil. Review* (1975), págs. 536-54.

y de las libertades, y de "un sentimiento de la propia dignidad"). Pero si lo peor que puede decirse de la justicia como imparcialidad es que pasa por alto la peculiar apelación del ascetismo, Rawls no tendrá que preocuparse mucho. Además, los ascetas no deben interesarse en consideraciones de justicia distributiva, sobre todo si no son envidiosos.

Pero, aunque es aceptable suponer que todo hombre racional desea tener más que menos de cualquier bien primario, no está tan claro que hombres racionales de diferentes convicciones respecto a los planes de vida muestren siquiera un acuerdo aproximado sobre el modo más adecuado de clasificar u ordenar diferentes grupos de bienes primarios. Rawls, en general, aborda este problema refiriéndose a "un índice de bienes primarios". Sin embargo, a excepción de la no muy útil unanimidad cuasi-ordenadora de la norma, no hay unos modos evidentes de establecer tal índice. Los "ingresos y riquezas" parecen bastante fáciles, porque, en una economía de mercado, el sistema de precios servirá como mecanismo indicador, pero, ¿cómo hemos de actuar respecto a las "oportunidades"?³⁷ Esta no es, de ningún modo, una cuestión trivial, según podemos ver imaginando cómo diferentes planes de vida originarían valoraciones o índices incompatibles de bienes primarios.

Comparemos dos hipotéticos "hombres representativos". El primero trata de maximizar su adquisición y disfrute de bienes de consumo, incluso al precio de una cierta artificialidad y estandarización de las incidencias de la vida. Está preparado para aceptar una disciplina laboral incluso rigurosa en su tarea, a trabajar largas horas y a ajustar sus patrones de consumo al catálogo siempre cambiante de bienes ofrecidos a la venta. El segundo valora la flexibilidad y la independencia en el trabajo, y un estilo más "natural" de consumo, basado en un mercado relativamente limitado, pero estable, de bienes muy conocidos. Los dos individuos tienen planes racionales de vida, pero sus índices de grupos de oportunidades, de riquezas y de ingresos, diferirán considerablemente.

Si una sociedad contiene dos grandes grupos de personas cuyos planes de vida muestran, en líneas generales, las formas características de los planes aquí esbozados, sus diferentes índices de bienes primarios se

³⁷ Para algunas observaciones esclarecedoras respecto al problema del índice, así como para una importante discusión de la regla maximin, ver Kenneth Arrow, "Some Ordinalist-Utilitarian Notes on Rawls' Theory of Justice", en *J. Phil.*, 70 (1973), págs. 241-63.

traducirán en políticas sociales competitivas, y, muy posiblemente, incompatibles. Optar por maximizar, simplemente el índice de bienes primarios de la persona representativa menos aventajada de la sociedad no *resolverá* el problema de los índices. Muy al contrario, lo *planteará* de un modo en que no lo planteará, por sí sola, una regla de maximización de ingresos y riquezas. Llegado el momento, Rawls trata de evitar este problema, que él conoce muy bien, sugiriendo que la distribución de bienes primarios no monetarios tiende a variar directamente con la de los ingresos y riquezas. Pero Rawls no presenta ninguna prueba de esta afirmación, y, a mi parecer, hay un límite para el número de supuestos *ad hoc* que un autor puede incorporar a un modelo formal destinado a generar un teorema importante³⁸.

Aquí, como en tantos otros sitios, vemos la concepción original de Rawls del problema de la elección operando por debajo de la superficie de su teoría final, mucho más elaborada. En su forma más sencilla (y más elegante), el problema de la elección era un problema de pura distribución de ingresos. El bien que se distribuía era mensurable, divisible, transferible, y, por consiguiente, manejable mediante el modelo formal de un juego del regateo. Pero en respuesta a críticas, dificultades e insuficiencias, Rawls adorna su retrato del problema de la elección hasta hacerlo totalmente inmanejable. Es bastante difícil, para un solo individuo con un plan racional de vida establecer un índice de un conjunto heterogéneo de bienes primarios para compararlo con otros conjuntos, sobre todo si no se le permite la fácil salida de "expresar", simplemente, una "preferencia" no razonada e inexplicable. Pero definir, aun del modo más tosco, un procedimiento de índice que ha de ser neutral entre planes de vida alternativos es, en mi opinión, sencillamente imposible. Cualquier intento tiene que hacer, inevitablemente, alguna apelación sustantiva a un tipo de planes de vida (o, quizás, a una clase de tipos de planes de vida), y, por consiguiente, elaborar unas preferencias normativas o culturales acerca de lo que se propone como procedimiento neutral de valoración.

Como he señalado, la concepción de Rawls de un plan racional o de vida se enmarca en una larga y honorable tradición filosófica. Su breve exposición de la concepción no puede ser más clara:

"La idea principal es que el bien de una persona está determinado por lo que para ella es el plan de vida más racional a largo plazo, en circunstancias razonables. Un hombre es feliz, en la medida en que tiene mayor o

³⁸ Ver Barber, "Justifying Justice", pág. 303.

menor éxito en llevar a cabo este plan. Para decirlo brevemente: el bien es la satisfacción del deseo racional. Hemos de suponer, entonces, que cada individuo tiene un plan racional de vida, diseñado según las condiciones a que se enfrenta, y también para permitir la satisfacción armónica de sus intereses. Sus actividades están programadas de modo que se puedan satisfacer sin interferencias los diversos deseos. Se llega a él rechazando otros planes que, o bien es menos probable que tengan éxito, o bien no hacen posible una realización tan amplia de fines. Dadas las alternativas disponibles, un plan racional es aquél que no puede mejorarse; no existe otro plan que, tomando todo en cuenta, pudiera ser preferible" (114-115).

Hay dos inconvenientes con esta noción de un plan de vida, que, en mi opinión, restan plausibilidad y aceptabilidad universal a la teoría de Rawls. El primero es que la concepción tiene unos límites excesivamente culturales, hasta el punto de que levanta, dentro de las exigencias pretendidamente formales de la posición original, ciertos supuestos no manifestados que dan expresión ideológica a una determinada configuración socio-económica y a un conjunto de intereses; el segundo es que, tal como Rawls lo concibe, el ideal de un plan racional de vida entra en conflicto con el carácter evolutivo, orgánico, de una personalidad humana sana.

Consideremos, primero, el tipo de sociedad humana en que fuese materialmente posible formular y mantener algo que respondiese a la descripción de Rawls de un plan racional de vida. La sociedad tendría que ser extremadamente estable —no en su posesión y reconocimiento de una concepción de la justicia, sino, simplemente, en la continuidad y predecibilidad de las líneas de una carrera, de las prácticas matrimoniales, de las disposiciones para la educación de los niños, etc. En una sociedad rápidamente cambiante, o en una sociedad cuyo futuro fuese materialmente imprevisible, los planes de vida se hundirían, convirtiéndose en planes de contingencia o también en métodos empíricos destinados a mantenerse atentos a cualquier incidencia que pudiera sobrevenir.

Incluso en una sociedad organizada y estable, el tipo de plan de vida que Rawls describe es más característico de las clases medias profesionales que de los hombres y mujeres de la clase obrera. En la América moderna, por ejemplo, los médicos tienden a seguir siendo médicos a lo largo de toda su vida de trabajo, y los profesores tienden a seguir siendo profesores; pero el obrero medio puede contar con que habrá de realizar trabajos que pertenecerán a muchas categorías completamente distintas, en el curso de su vida laboral. Ahora bien: Rawls sabe

esto, naturalmente, y su discusión está jalonada de matices, cautelas y reconocimientos de la incertidumbre de la existencia humana. Enfrentado con los hechos del cambio medio de trabajo, por ejemplo, un joven americano podría llegar a la conclusión de que una educación liberal general, adecuada a las demandas de flexibilidad y repetida "readaptación", sería, "tomándolo todo en cuenta", la primera etapa más racional de un plan de vida.

Pero llega un momento en que los matices destruyen la coherencia de una concepción. Si prescindimos un tanto de los ornamentos del lenguaje de Rawls, a veces excesivamente protector, e intentamos formarnos una imagen del tipo de persona que se ajustaría a sus descripciones, aparece muy claramente un hombre profesional (el libro está sobrecargado de un lenguaje de orientación masculina), lanzado a una carrera, viviendo en un ambiente político, social y económico estable, en el que pueden adoptarse decisiones razonadas acerca de cuestiones a largo plazo como los seguros de vida, la localización residencial, la escolarización de los niños y la jubilación. La orientación temporal se dirige esencialmente hacia el futuro, más que hacia el pasado, pero hacia el futuro como conjunto, no hacia un determinado momento dramático o exaltado del futuro. En resumen, lo que vemos es exactamente lo que Karl Mannheim caracterizaba, en su brillante discusión de la estructura ideológica de la conciencia temporal, como la mentalidad utópica liberal humanitaria¹⁹.

El segundo problema es que, como todas las ideologías, esta concepción de los planes racionales de vida es una *tergiversación* de la realidad de la experiencia humana, no una representación correcta de un estilo especial de experiencia humana. En consecuencia, además de ser especial más que general, es también distorsionada más que exacta. El modelo de racionalidad aquí invocado por Rawls es un modelo apropiado a una empresa, más que a un ser humano individual. En cierto sentido, Rawls adapta a sus propios usos un modelo de argumentación desplegado por Platón en el *Gorgias*. Platón comienza con la noción ordinaria de una *techné*, o "arte", y la condición, o "virtud", del "artista", en virtud de la cual puede desempeñar bien su propia función como artista. Pregunta luego cuál es la función propia o *techné* del alma, concluye que es "la de vivir", e infiere que debe haber alguna sana condición del alma, alguna virtud que le permita vivir bien — es decir, desempeñar su propia función, adecuadamente. Así también, Rawls

¹⁹ Karl Mannheim, *Ideología y utopía*, parte IV, F.C.E. México, 1941.

comienza con la noción de racionalidad prudencial apropiada a la actividad económica, y con la noción asociada de que una empresa racional tendrá un plan a largo plazo de maximización de la ganancia, y no simplemente una tendencia establecida a beneficiarse de todas las oportunidades generadoras de ganancia que se presenten cada día. Empleando como análisis de la racionalidad individual los modelos formales que se han desarrollado en una teoría de la actividad económica, Rawls trata el hecho de vivir una vida como análogo a la dirección de una empresa. Una vida racionalmente vivida estará guiada por un plan de gran autonomía, y una vida, como una empresa, estará funcionando bien, cuando el plan está "más o menos afortunadamente en vías" de realización.

Hay, naturalmente, problemas con la noción económica de un plan racional de maximización de la ganancia, siendo uno de los más complejos el del espacio de tiempo a lo largo del cual han de recopilarse los resultados para calcular los rendimientos. A la luz de estas observaciones, no es de extrañar que Rawls trate la relación entre padres e hijos como inextricablemente unida a una elección apropiada de una tasa social de ahorros.

Pero el hecho de vivir una vida no es, en absoluto, como la administración de una empresa. Una empresa es una persona legal, pero no una criatura orgánica, natural, viva. Un ser humano tiene una infancia, una niñez, una adolescencia, una edad adulta joven, una edad adulta madura, y una vejez. Lo que es "racional" en una etapa del ciclo de la vida es irracional, insano, en otra. Es lógico que un joven o una joven sueñen con grandes triunfos, con un futuro ilimitado, con un tiempo sin fin. En un hombre o en una mujer de edad madura, esos mismos sueños son una huida de la realidad, una fantasía contraria a lo real. Un adolescente que formula un plan de vida, completado con fondos de reserva, con asignaciones para contingencias y con un persistente interés por un sólido plan de pensiones, perdería, casi seguramente, una gran parte de la alegría y de las satisfacciones que la vida nos ofrece. Cuando leo que Immanuel Kant, en dos ocasiones, dedicó tanto tiempo a calcular los beneficios y las cargas del matrimonio, que el objeto de sus posibles afectos adoptó otras disposiciones matrimoniales, no creo que estuviese proyectando racionalmente un prudente plan de vida; lo que deduzco es que no tenía madera de casado, y que, evidentemente, se encontraba mejor soltero.

Michael Oakshott, el notable pensador y político conservador inglés, ha captado perfectamente la singular noción de "razón" que

subyace en la noción de un plan racional de vida en su brillante ensayo, "Rationalism in Politics". Las siguientes líneas dicen, mejor de lo que yo pudiera hacerlo, cuáles son los errores de la concepción de Rawls:

"El espíritu de los racionalistas... nos impresiona, en el mejor de los casos, como un espíritu de fino temple, un instrumento neutral, bien preparado, más que como espíritu ilustrado. Intelectualmente, su ambición no consiste tanto en compartir la experiencia de la especie como en ser un hombre que se hizo a sí mismo y en poder demostrarlo. Y esto da a sus actividades intelectuales y prácticas un carácter casi preternatural de deliberación y de auto-conciencia, privándolas de todo elemento de pasividad, eliminando de ellas todo sentido de ritmo y de continuidad, y disolviéndolas en una sucesión de momentos climatéricos, cada uno de los cuales ha de ser superado por un *tour de raison*. Su espíritu no tiene atmósfera, ni cambios de estación y temperatura; sus procesos intelectuales, en la medida de lo posible, están aislados de toda influencia externa y avanzan en el vacío... Con una fantasía casi poética, se esfuerza por vivir cada día como si fuese su primer día, y cree que formarse un hábito es fracasar."⁴⁰

No puedo imaginar una caracterización más perfecta de las partes en la posición original, ni un retrato menos persuasivo de la verdadera racionalidad.

⁴⁰ Michael Oakshott, *Rationalism in Politics* (Basic Books, 1962), págs. 2-3.

UN ANALISIS FORMAL DEL JUEGO DEL REGATEO

UNA VEZ MÁS, será conveniente abordar el problema por etapas. En primer lugar, analizaremos el juego descrito por Rawls en la primera versión de su teoría, tal como se expone en "La justicia como imparcialidad". Después, examinaremos brevemente una calificación, agregada por Rawls a ese ensayo, que altera el juego, pero que no altera de un modo importante nuestro análisis. En tercer lugar, tenderemos el velo de la ignorancia y reconstruiremos el juego consecuentemente, dejando a un lado, sin embargo, los problemas especiales suscitados por el hecho de que los jugadores bajo el velo no saben en qué tiempo ni en qué espacio están. Una vez obtenida una idea clara de esa forma del juego, complicaremos las cosas para considerar este último fragmento de ignorancia en la posición original, y para ver luego lo que hemos alcanzado por la vía de un juego del regateo.

Aclarada ya la estructura del juego, podemos luego confrontar el complejo y polémico uso que Rawls hace de la regla maximin de elección bajo incertidumbre. Ofreceré unas observaciones ligeramente técnicas sobre el tema, tomadas, en parte, de Luce y Raiffa, de John Harsanyi y de otros, con el propósito de poner en cuestión ese uso. Por último, a manera de conclusión, diré algo acerca de la noción de la justicia procesal pura.

ESTRUCTURA DEL JUEGO

El Juego en "La justicia como imparcialidad"

Aceptaré, por razones de simplicidad, que el juego del regateo de Rawls tiene que construirse como un juego no-cooperativo, en el que no hay comunicación previa⁴¹. A cada jugador se le permite, pues, una sola jugada, que consiste en escribir su propuesta para los principios de

⁴¹ Las razones de este supuesto se exponen en el capítulo V.

la justicia en un trozo de papel y en entregarlo al árbitro. Para dar mayor relieve a la teoría estándar del juego, supondremos que no hay más que un número elevado, pero finito, de principios diferentes que pueden ser propuestos, y que a cada jugador se le entrega una relación de ellos, n en total. Cada jugador dispone, pues, exactamente, de $n + 1$ estrategias: puede escribir cualquiera de los n principios de la relación, o puede dejar en blanco su papel. Esta última alternativa equivale a optar por el estado de naturaleza que predominaba antes del comienzo del juego⁴².

Suponiendo que no hay más que un número elevado, pero finito, de jugadores, por ejemplo, m , se sigue que hay $(n + 1)^m$ combinaciones diferentes de estrategias que pueden ser entregadas al árbitro, y cada una de ellas estará asociada, por las reglas del juego, con un resultado. En el caso más simple, en el que sólo hay un principio posible de la justicia y dos jugadores, cada uno de los jugadores tendrá dos estrategias (proponer ese principio o dejar en blanco su papel), y habrá cuatro resultados —lo que nos da la conocida matriz de resultados dos-por-dos, que tan frecuentemente se ve en los textos de la teoría del juego. Como las reglas del juego dicen que se adoptará un principio cuando sea elegido por unanimidad —y sólo entonces—, la matriz de resultados de este juego sencillo se asemejará a la figura 1.

		Jugador 2	
		Principio 1	Dejar Papel Blanco
Jugador 1	Principio 1	Se adopta Principio 1	Estado de Naturaleza
	Dejar Papel Blanco	Estado de Naturaleza	Estado de Naturaleza

Figura 1

⁴² En esta versión del juego, hay exactamente tantas estrategias para cada jugador como principios hay, más la de dejar en blanco el papel. En cualquier forma de la versión cooperativa, incluso en una que sólo permita una única vuelta al correo, el número de estrategias se eleva astronómicamente. Para una definición formal de una estrategia, ver Luce y Raiffa, *Games and Decisions*, págs. 51-52. No trataré de tomar en cuenta la posibilidad de estrategias mixtas, que realmente enreda las cosas de un modo tremendo.

Cuando hay dos jugadores y más principios disponibles, la matriz seguirá siendo bidimensional, pero será mayor. Con cuatro principios, tendremos una matriz de resultados cinco-por-cinco, semejante a la figura 2.

		Jugador 2				
		Princ. 1	Princ. 2	Princ. 3	Princ. 4	Blanco
Jugador 1	Princ. 1	Princ. 1 adoptado	Estado de Naturaleza	Estado de Naturaleza	Estado de Naturaleza	Estado de Naturaleza
	Princ. 2	Estado de Naturaleza	Princ. 2 adoptado	Estado de Naturaleza	Estado de Naturaleza	Estado de Naturaleza
	Princ. 3	Estado de Naturaleza	Estado de Naturaleza	Princ. 3 adoptado	Estado de Naturaleza	Estado de Naturaleza
	Princ. 4	Estado de Naturaleza	Estado de Naturaleza	Estado de Naturaleza	Princ. 4 adoptado	Estado de Naturaleza
	Blanco	Estado de Naturaleza	Estado de Naturaleza	Estado de Naturaleza	Estado de Naturaleza	Estado de Naturaleza

Figura 2

Para m jugadores, requeriremos una formación m -dimensional, cada una de cuyas hileras será de una longitud de $n + 1$ términos. En todas las casillas de la formación se leerá "Estado de Naturaleza", excepto en las que corresponden a la unánime coordinación sobre uno u otro de los n principios disponibles. Hay, pues, $n + 1$ resultados distintos, y una pequeña imaginación geométrica descubrirá que componen una diagonal de la formación m -dimensional, que se extiende, por así decirlo, de noroeste a sudeste.

El paso siguiente es que cada jugador calcule el valor que para él tiene cada uno de los resultados posibles, de modo que podemos construir una "matriz de remuneraciones". La matriz de remuneraciones tiene la misma estructura que la matriz de resultados, pero donde la matriz de resultados registra un resultado ("Principio 1 adoptado", "Principio 9 adoptado", "Estado de Naturaleza", etc.), la matriz de remuneraciones registrará un ordenado m -tuplo de números que representarán el valor de ese resultado para cada uno de los m jugadores.

Para empezar, cada jugador conoce, probablemente, qué valor asigna al estado de naturaleza en que él se encontraba antes de entrar en el juego del regateo, de modo que inserta ese valor donde lo considera apropiado. Cada jugador conoce también las evaluaciones de cada uno de los otros jugadores, y a nadie está permitido tergiversar sus preferencias (todos éstos son supuestos desoladoramente inadmisibles, supuestos de la teoría standard del juego). Pero las cosas se hacen más difíciles cuando hay que asignar los valores a los resultados no-estado-de-naturaleza. Lo que cada jugador tiene que hacer es efectuar un cálculo sumamente complicado de la utilidad que se espera, en el que estime el sistema de prácticas con la distribución asociada que es probable que se produzca en su sociedad, si ésta se gobierna por el principio que se considera; que estime las probabilidades que asignará a las diversas posiciones representativas en ese sistema; que calcule el valor que para él tiene la ocupación de aquellas posiciones, debidamente descontado por sus probabilidades; y, finalmente, dadas incertidumbres de la vida como el desempleo, perjuicios industriales, etc., que realice una gran suma de todos los resultados posibles, multiplicados por sus valores y descontados por sus probabilidades. Inserta ese número en su muesa, en el m -tuplo adecuado. Hace esto con cada principio, y, una vez más, suponemos que conoce, o que se le dice, el valor que cada jugador coloca en cada muesa. Cuando ya cada jugador ha hecho sus cálculos y todos tienen una matriz de remuneración correcta y totalmente rellenada al lado de la relación de principios disponibles y su papeleta, entonces podemos iniciar el juego⁴¹.

Llegado a este punto, Rawls propone que cada jugador elija una estrategia por medio de la regla maximin. Esta regla se explica habitualmente en relación con juegos de dos personas, que pueden representarse, gráficamente, mediante una formación bidimensional de hileras y columnas. Entonces, puede hablarse del "mínimo ingreso en una hilera" o del "máximo ingreso en una columna", etc. Sin embargo, en un juego de m -personas, las remuneraciones están representadas por una formación m -dimensional, y los resultados correspondientes a una determinada estrategia elegida por un jugador constan de una formación $(m - 1)$ -dimensional, y no de una hilera o columna. (Una hilera, natu-

⁴¹ Nada se ha dicho acerca de la comparabilidad interpersonal de los números en la matriz de remuneraciones, pero, a la vista del proceso por el que se derivan, está claro que son fundamentalmente importantes. Sin embargo, estas cuestiones no desempeñan un papel en nuestra discusión, hasta más adelante, en esta sección.

ralmente, no es más que una formación uni-dimensional, es decir, una formación $[2 - 1]$ -dimensional). Aun a riesgo de crear alguna confusión, hablaré, en lo que sigue, de formaciones multidimensionales, no de hileras o columnas. Pero los ejemplos gráficos mostrarán, en su totalidad, la forma de las conocidas matrices bidimensionales.

De todos modos, un jugador examina primero la formación $(m - 1)$ -dimensional que representa todas las diferentes remuneraciones que él puede obtener como consecuencia de la adopción de una determinada estrategia (según lo que los otros jugadores hagan), y sitúa la remuneración mínima en la formación. Este es su "nivel de seguridad". Si tuviera que elegir esa estrategia, esto es lo peor que podría ocurrirle. Después estudia el nivel de seguridad de cada una de las estrategias de que dispone, y selecciona la estrategia con el más alto nivel de seguridad — la estrategia con la máxima remuneración mínima. Al realizar este cálculo, ignora todas las demás remuneraciones, cualesquiera que sean la magnitud y la distribución de las mismas. Tampoco hace intento alguno de estimar la probabilidad de las elecciones de los otros jugadores. Si hay varias estrategias con el mismo nivel máximo de seguridad, y si ninguna de ellas puede dejar de ser considerada en virtud de lo que se conoce como "dominación débil"⁴⁴, entonces todo el conjunto de estrategias con el máximo nivel de seguridad es considerado como la elección "correcta" para el jugador, y éste, sencillamente, selecciona una, al azar.

En el juego determinado que estamos examinando, la remuneración de un jugador será la misma para todas las casillas de una formación dada, j (concretamente, el valor que sitúa en el estado de naturaleza), excepto para la casilla correspondiente al resultado "Principio j adoptado". Hay tres posibilidades: la adopción del principio j puede, según el cálculo del jugador, ser para éste más valiosa que el estado de naturaleza, menos valiosa que el estado de naturaleza, o del mismo valor que el estado de naturaleza. Si la adopción del principio j es para él más valiosa que el estado de naturaleza, entonces su nivel de seguridad para esa formación será el valor de su estado de naturaleza. Si el principio j tiene el mismo valor para él que el estado de naturaleza, el valor

⁴⁴ Una estrategia "domina débilmente" a otra, si la primera resulta, por lo menos, tan bien como la segunda, independientemente de lo que los otros jugadores elijan, y, en realidad, resulta mejor, por lo menos, para una de las elecciones de los otros jugadores. Evidentemente, si un jugador dispone de dos de esas estrategias, nunca elegirá la segunda, porque la primera *debe* ser, por lo menos, tan buena para él, y *puede* ser mejor.

del estado de naturaleza será su nivel de seguridad. Pero si el principio *j* es para él menos valioso que el estado de naturaleza, entonces ese valor menor será su nivel de seguridad.

Consideremos ahora que una formación determinada, la correspondiente a la opción del estado de naturaleza, debe tener el valor del estado de naturaleza como su nivel de seguridad. Más aun: ninguna formación puede tener más que el valor del estado de naturaleza como su nivel de seguridad, porque todas las formaciones tienen muchas casillas de estado de naturaleza. Por lo tanto, el máximo nivel de seguridad de la matriz, para cada jugador, es el nivel de estado de naturaleza. Hay exactamente un conjunto de circunstancias, y sólo uno, en el que se puede confiar que un jugador racional, a la hora de elegir su estrategia mediante la regla maximin, seleccionará como su estrategia el principio de diferencia. Si *cualquier otro principio* ofrece al jugador una remuneración más baja, en el caso de que lo adopte, que el estado de naturaleza, y el principio de diferencia le ofrece más que el estado de naturaleza, entonces elegirá el principio de diferencia. Las consideraciones del nivel de seguridad eliminarán todas las estrategias, excepto el principio de diferencia y la opción del estado de naturaleza, y las consideraciones de dominación débil eliminarán la opción del estado de naturaleza. Pero, si hay *dos* principios (teniendo en cuenta las transferencias y las compensaciones) que, si se adoptan, le ofrecen más que el estado de naturaleza, entonces la maximin le inducirá a sortear entre el conjunto de esos principios, y con *m* jugadores sorteando de este modo, la probabilidad de coordinación será extremadamente exigua.

No hay manera de determinar, en general, si habrá algún principio —o muchos principios— cuya adopción ofrezca a un jugador dado más de lo que él puede obtener del estado de naturaleza, pero creo que es justo decir que Rawls piensa que habrá, por lo menos, varios, y, de ser así, entonces su teorema no es válido.

El Juego en "La justicia como imparcialidad" con pesimismo agregado

Hasta ahora, hemos venido analizando la primera forma del modelo de Rawls sobre el supuesto de que los jugadores rellenan la matriz de las remuneraciones llevando a cabo los cálculos de utilidad previstos. Pero, en "La justicia como imparcialidad", parece que Rawls sugiere que los jugadores proceden de un modo diferente, y mucho más pesimista. Los principios propuestos por los jugadores "expresarán las condiciones en que cada uno de ellos está dispuesto a ver limitados sus

intereses en el proyecto de las prácticas, dados los intereses competitivos de los otros, en el supuesto de que los intereses de los otros sean limitados también. Las restricciones que así surgirían podrían ser consideradas como las que una persona tendría en cuenta si estuviese proyectando una práctica, en la que su lugar hubiera de serle asignado por su enemigo" (138-139).

El lenguaje, en especial la expresión "está dispuesto", sugiere que Rawls concibe esta propuesta como una forma de maximin, pero, en realidad, es una propuesta para construir la matriz de remuneraciones a partir de la matriz de resultados, y no una propuesta para elegir una estrategia, una vez que se haya completado la matriz de remuneraciones. Como no se han hecho supuestos de ningún tipo acerca de las funciones de utilidad de los diversos jugadores, esta nueva restricción no bastará para que la matriz de remuneraciones se haga simétrica⁴¹. Para algunos jugadores, reducirá el volumen del conjunto de estrategias sobre las cuales echan suertes (en términos generales, reduce el conjunto, para un jugador cuyas facultades naturales, o de cualquier otro tipo, le dan prometedores resultados de los diversos cálculos de utilidad previstos, efectuados para construir la matriz de remuneraciones); pero la forma del problema continúa inalterada. Ni siquiera una doble aplicación de la maximin, primero en la construcción de la matriz de remuneraciones y luego en la elección de estrategia, facilitará la "solución" que Rawls busca.

El Juego bajo el Velo de la Ignorancia: primera aproximación

Cuando se tiende el velo de la ignorancia, el juego cambia en un buen número de aspectos sumamente importantes. Los jugadores ya no saben quiénes son, dónde se hallan situados ellos o sus sociedades en el tiempo histórico o en el espacio, cuáles son sus funciones de utilidad, ni siquiera cuántos jugadores hay en un momento dado del tiempo real. En tales circunstancias, resulta un poco difícil, para decirlo suavemente, la construcción de una matriz de remuneraciones. Hay, sin embargo, una ventaja compensadora, a la que Rawls concede mucha importancia: al desechar complicaciones tales como los elementos de oportunidad o de azar en la determinación de los valores de las remuneraciones, la ignorancia de los jugadores destruye todas las bases sobre las cuales éstos pudieran asignar diferentes valores a un resulta-

⁴¹ Esto es, los *m*-tuplos, en general, no serán cadenas de números idénticos.

do dado. En consecuencia, la matriz de las remuneraciones, si se construyese, sería *simétrica*.

Los jugadores todavía no tienen idea alguna de lo que necesitan, y, en consecuencia, no tienen modo alguno de saber en qué medida les satisfarán los diversos estados de los asuntos. Rawls supone, por lo tanto, que todos los jugadores tienen un "plan de vida"; que algún subconjunto identificable de las cosas distribuibles en una sociedad es especialmente útil para proseguir cualquier plan racional de vida que una persona pueda adoptar, y, en consecuencia, puede recibir el nombre de "bienes primarios"; que ningún otro tipo de cosas distribuibles tiene una excesiva importancia para ningún plan racional de vida, de modo que la utilidad de un jugador dependerá solamente de lo que obtenga de esos bienes primarios; y, naturalmente, que todos los jugadores necesitan más bienes primarios que menos, de modo que tienen una utilidad marginal positiva (sobre algún intervalo realista) para un índice adecuado de bienes primarios. Las condiciones de la ignorancia requieren también, como una simplificación altamente significativa, que se imagine que todos los jugadores tienen exactamente la misma función de utilidad.

El primer problema en la construcción de la matriz de remuneraciones es la selección de un valor para los resultados del estado de naturaleza. Aun cuando se suponga que los jugadores viven, antes del regateo, en las "condiciones de justicia", una de las cuales es una igualdad aproximada en el poder, debemos suponer que hay una diferencia muy importante entre la condición de los individuos representativos peor y mejor acomodados en el estado de naturaleza. El carácter del razonamiento de Rawls aclara que sólo hay dos valores posibles para las casillas del estado de naturaleza: o el del individuo peor acomodado en el estado de naturaleza o el del mejor⁴⁶. Llegados a este punto, suponemos, provisionalmente, que los jugadores saben en *qué* sociedad viven, pero no *quiénes* son en la sociedad. Es de recordar que los estados de naturaleza no son muchachas salvajes en taparrabo, o guerras de todos contra todos, sino sencillamente, situaciones sociales en las que no se ha llegado a ningún acuerdo acerca de unos principios de la justicia determinados.

⁴⁶ Podría parecer atractivo, a manera de compromiso, algún tipo de promedio, pero, si Rawls permite que sus jugadores se comprometan en un promedio a fin de llegar a un valor para el estado de naturaleza, le será difícil negarles el mismo recurso a la hora de evaluar los resultados para los diversos principios, y por ese camino se encuentra el utilitarismo y Dios sabe qué otras cosas.

Ahora bien: la idea subyacente en el argumento de Rawls consiste en que los participantes en el juego del regateo deben ser prudentes, conservadores, incluso pesimistas en sus cálculos. No deben lanzarse de cabeza a un acuerdo que después habrán de lamentar. Si un jugador supone que él es una de las personas peor acomodadas en su estado de naturaleza, estará excesivamente deseoso de llegar a un acuerdo, porque todo principio cuya adopción prometa a la persona peor acomodada más de lo que la persona peor acomodada consigue en el estado de naturaleza resultará atractivo para él. Cuando recobra el pleno conocimiento de sí mismo, puede descubrir que él *no* era uno de los peor acomodados en el estado de naturaleza, y que su acuerdo ha rebajado, en realidad, su resultado. Así, el *pesimismo* le induce a suponer que está mejor en el estado de naturaleza. El mismo pesimismo, naturalmente, le induce también a suponer que estará peor bajo la acción de los principios acordados. Así, parecería que el simple conjunto de normas siguiente encierra la concepción de Rawls del modo racional de construir la matriz de remuneraciones:

1. Para cada resultado de estado de naturaleza, asígnese el valor de ese estado de naturaleza para el individuo representativo mejor acomodado en él.

2. Para cada resultado de adopción-de-un-principio, asígnese el valor de la adopción de ese principio para el individuo representativo peor acomodado bajo él.

Esta propuesta sólo plantea un problema. Si suponemos, como parece razonable, que el individuo representativo *peor acomodado* bajo cualquiera de los principios propuestos no estará tan bien como el individuo representativo *mejor acomodado* en el estado de naturaleza (después de todo, las posibilidades de cooperación social tienen sus límites), entonces la opción del estado-de-naturaleza dominará débilmente cualquier otra estrategia, y la consecuencia del juego será un retorno al estado de naturaleza.

Si los jugadores hacen el supuesto *optimista* de que son los individuos peor acomodados en el estado de naturaleza, y lo unen al supuesto pesimista de que serán los peor acomodados bajo cualquiera de los principios, entonces tendremos una matriz de remuneraciones de cierto interés. La regla maximin de elección seguirá sin permitirnos alcanzar una determinada solución, por todas las razones analizadas más arriba, pero con un cierto número de nuevos supuestos y complicaciones, podremos llegar, realmente, al principio de diferencia de Rawls. Las matrices de las figuras 3, 4 y 5, para dos jugadores y dos principios más

el estado de naturaleza, aclararán la situación. Los números, como de costumbre, sólo tienen un carácter ilustrativo.

		<i>Jugador 2</i>		
		Principio 1	Principio 2	Blanco
<i>Jugador 1</i>	Principio 1	11,17	10,15	10,15
	Principio 2	10,15	18,13	10,15
	Blanco	10,15	10,15	10,15

Figura 3. La matriz "verdadera"

		<i>Jugador 2</i>		
		Principio 1	Principio 2	Blanco
<i>Jugador 1</i>	Principio 1	11,11	15,15	15,15
	Principio 2	15,15	13,13	15,15
	Blanco	15,15	15,15	15,15

Figura 4. La matriz "pesimista"

		<i>Jugador 2</i>		
		Principio 1	Principio 2	Blanco
<i>Jugador 1</i>	Principio 1	11,11	10,10	10,10
	Principio 2	10,10	13,13	10,10
	Blanco	10,10	10,10	10,10

Figura 5. La matriz "optimista"

La figura 3 muestra los verdaderos valores, para los jugadores 1 y 2, del estado de naturaleza y de la adopción de los principios 1 y 2 (dejando a un lado todas las cuestiones de comparaciones interpersonales de utilidad, etc.). La figura 4 es la matriz de remuneraciones que se construiría mediante la adopción de un enfoque pesimista del proceso de regateo: cada jugador supone que está mejor en el estado de naturaleza y que estará peor bajo cualquier principio adoptado. Está muy claro que la tercera estrategia de cada jugador (la de dejar en blanco) domina débilmente a las otras dos, y la consecuencia será una reversión al estado de naturaleza. La figura 5 es la matriz de remuneraciones construida mediante la combinación del optimismo acerca del proceso de regateo en general (es decir, que uno tiene mucho que ganar con el proceso), con el pesimismo acerca de las consecuencias del proceso (que uno estará peor bajo cualquier conjunto de principios que se adopte). No hay muy buenas razones para optar por este punto de vista, pero proporciona un juego manejable. Las estrategias 1 y 2 dominan débilmente a la estrategia 3 para cada jugador, y así el juego se reduce a un juego de dos-por-dos. La maximin seguirá dando ambas estrategias como la solución establecida. Pero la simetría nos induce a creer que los jugadores cooperarán de acuerdo con el principio 2 si se les permite comunicarse, o se coordinarán según el principio 2, si no se les permite.

¿Podemos, entonces, llegar finalmente a la conclusión de que hemos encontrado un modo de construir el juego que proporcione el principio de diferencia como solución? Como el principio de diferencia, por definición, maximiza la remuneración para el individuo representativo peor acomodado (y luego, lexicográficamente, para cada individuo representativo menos necesitado), la casilla para la adopción del principio de diferencia, dada la simetría, será, seguramente, el resultado de la cooperación o de la coordinación.

En absoluto. Rawls tiene que establecer un nuevo supuesto, antes de que pueda adoptar el análisis aquí presentado (o su propio análisis, por lo demás) y declarar el principio de diferencia como solución del juego. El supuesto puede parecer trivial, y su mención puede parecer torpemente inoportuna, pero, por razones que trataré de explicar en la quinta parte, creo que la facilidad con que Rawls establece el supuesto nos dice mucho acerca de lo que hay de erróneo en su aproximación al tema de la filosofía social.

En resumen, debemos suponer que, cuando se adopta un principio, es seguro que la sociedad que lo ha adoptado llegará, con el tiempo, a

establecerlo como norma. Rawls ha introducido este supuesto extraordinariamente poderoso en su argumentación, ya en las páginas iniciales, mediante su definición de una "sociedad bien ordenada" como una sociedad en la que "(1) cada cual acepta y sabe que los otros aceptan los mismos principios de justicia, y (2) *las instituciones sociales básicas satisfacen generalmente* estos principios y se sabe generalmente que lo hacen (pág. 21, la cursiva es mía).

Rawls es libre de limitarse a una consideración de las sociedades bien ordenadas, como ya he señalado, pero *no* es libre de suponer sin pruebas que cualquier principio determinado, como el principio de diferencia, puede, en realidad, *ser* el principio de una sociedad bien ordenada. Un conocimiento de los hechos básicos de la sociedad podría revelar que, cuando una sociedad adopta conscientemente el principio de diferencia, no puede, a pesar de sus mejores esfuerzos, maximizar el índice de bienes primarios asignados al individuo representativo peor acomodado. Los economistas y los sociólogos, desde Adam Smith hasta el presente, nos han advertido acerca de las inesperadas consecuencias de las políticas sociales públicamente más vivas y moralmente más admirables. Elevad el salario mínimo —asegura Milton Friedman—, y aumentaréis el desempleo, al impulsar a los capitalistas a introducir técnicas de ahorro de fuerza de trabajo que, con un salario más bajo, no serían rentables. Apretad los tornillos en lo que se refiere a la entrada de heroína en el país, y provocaréis una oleada de crímenes, pues la elevación de precio en el mercado negro empuja a los adictos a robos cada vez más violentos. Al dejar a un lado tales consideraciones, Rawls niega, sencillamente, el carácter social de la sociedad.

El Juego bajo el Velo de la Ignorancia: segunda aproximación

En la versión oficial de la situación original, los jugadores ni siquiera saben de qué sociedad real proceden. No tienen idea alguna acerca de cuál es su nivel de tecnología, ni de qué etapa del desarrollo socio-político ha alcanzado. Pero conocen la verdad de un gran número de proposiciones condicionales, de las que puede imaginarse que tienen la forma siguiente: "Si estamos en una sociedad y en una economía de tipo t , la adopción del principio j conducirá, suponiendo que la sociedad esté bien ordenada, a un patrón y a un nivel de distribución de bienes primarios, d , en los que la población se dividirá en categorías de ingresos, o en clases sociales, o en individuos x , y , z , etc." (Ya he indicado por qué creo que sería imposible que los jugadores poseyesen sólo in-

formación de este tipo general.) De ello se sigue que no habrá un solo estado de naturaleza al que los jugadores retrocedan en ausencia de un acuerdo, sino tantos estados de naturaleza diferentes como diferentes realidades sociales posibles haya. Y a cada principio de la lista de principios de justicia disponibles estará asociado no *un solo* resultado en el caso de que se adopte, sino una gama de resultados, uno para cada diferente realidad social en la que pueda ser puesto en práctica.

La forma más natural de representar esta complicación adicional es la de construir una matriz de resultados $m + 1$ dimensional, en la que la "realidad" sea tratada como el jugador $m + 1$.⁴⁷ Podemos imaginar que, mientras cada jugador está seleccionando una estrategia entre los n principios y la opción del estado de naturaleza, la "realidad" está seleccionando, independientemente, un estado socio-económico real de los asuntos en el que el principio será puesto en práctica. Los jugadores saben qué opciones tiene la realidad, y saben cuál será el resultado de combinar cualquiera de esas opciones con uno de los principios disponibles (esto es lo que significa, en mi opinión, conocer los hechos básicos de la sociedad, etc.). En consecuencia, podrán construir la matriz de resultados.

En la construcción de la matriz de remuneraciones, los jugadores tienen que ser, evidentemente, muy optimistas acerca del juego del regateo en cuanto tal — lo que significa que tienen que asignar a cada posible resultado de estado de naturaleza el valor de ese resultado para el individuo peor acomodado bajo él. Pero, como tenemos una dimensión de la matriz a lo largo de la cual varía la realidad social, habrá tantos valores diferentes de ese tipo, como realidades sociales posibles diferentes haya.

Rawls sostiene que es racional que los jugadores sean completamente pesimistas (o cautos) en la valoración de los resultados correspondientes a la adopción de los diversos principios en las diversas realidades sociales posibles. Esto significa asignar a cada uno de esos resultados el valor de esa realidad social, organizada según el principio en cuestión, para el individuo representativo peor acomodado. Tendremos que examinar las razones de Rawls para proponer una regla tan

⁴⁷ En las discusiones estándar de los problemas de decisión bajo incertidumbre, el jugador extra se denomina "naturaleza", y los juegos de este tipo se llaman "juegos contra la naturaleza". En este caso, sin embargo, no es la naturaleza de la naturaleza la que está en cuestión, sino la naturaleza de la sociedad, por lo que "realidad" es un término mejor. Cf. Luce y Raiffa, *Games and Decisions*, capítulo 13, donde se discute ampliamente el caso de la elección individual bajo la incertidumbre.

pesimista para la construcción de la matriz de remuneraciones.

Finalmente, con la matriz de remuneraciones establecida, Rawls sostiene que la regla racional de elección es la maximin. (Como la matriz de remuneraciones, bajo el velo de la ignorancia, es simétrica, el problema de la elección es, naturalmente, el mismo para cada jugador.) Ahora, sin embargo, la maximin tiene una significación nueva, y doblemente pesimista. A medida que un jugador considera cada estrategia sucesivamente, examina el ordenamiento de los posibles resultados compatibles con esa elección de estrategia. Hay $(n + 1)$ estrategias disponibles para los demás $(m - 1)$ jugadores, y la "realidad" tiene un enorme número, S , de posibles situaciones sociales a su disposición, de modo que el ordenamiento constará de $S(n + 1)^{m-1}$ remuneraciones. Pero este ordenamiento *no* constará de un único valor de estado-de-naturaleza en todos los lugares menos en uno. La actual distribución de valores será un poco más complicada. Para cada uno de los S estados de naturaleza, hay $(n + 1)^{m-1}$ resultados, correspondientes a todas las formas en que los otros $m - 1$ jugadores pueden elegir entre las $n + 1$ estrategias de que disponen. Algunas de éstas, n en número, serán coordinaciones según principio de justicia, de modo que para cada posible estado de naturaleza habrá $[(n + 1)^{m-1} - n]$ resultados idénticos, cada uno de ellos con un valor igual a la remuneración en ese estado de naturaleza para el individuo y n resultados diferentes, uno por el valor para el individuo peor acomodado bajo cada uno de los n principios, tal como se halla establecido en ese estado de naturaleza. Como hay S estados de naturaleza (¡un número muy elevado!), habrá $S[(n + 1)^{m-1} - n] + Sn$ resultados totales, lo que, como se ve, es igual a $S(n + 1)^{m-1}$; y habrá $S + Sn = S(n + 1)$ resultados distintos en todo el ordenamiento.

Cuando un jugador observa las $S(n + 1)^{m-1}$ casillas correspondientes a una determinada elección de estrategia, ve todas las $S(n + 1)$ casillas distintas y selecciona la más pequeña absolutamente. Podemos suponer que éste será el valor para el individuo peor acomodado del peor estado de naturaleza⁴⁴. Como este valor se presenta en cada uno de los $n + 1$ ordenamientos a las elecciones de estrategia del jugador, será su máximo nivel de seguridad. Al igual que antes, la dominación débil eliminará la opción del estado de naturaleza, y la cooperación o la coordinación ofrecerán una razonable esperanza de llegar

⁴⁴ Puede ser, pero no es necesario que sea, el estado de naturaleza que opera al nivel tecnológico más bajo, al que se pretende aplicar el principio de diferencia. En todo caso, podemos estar seguros de que es bastante malo.

al principio de diferencia como a la elección unánime de estrategia.

El propósito de este revoltijo técnico ha sido el de exponer los supuestos que subyacen en la argumentación informal y más bien abocetada de Rawls. Algunos puntos están claros ya. Primero: Rawls supone que los jugadores podrán, en la construcción de la matriz de remuneraciones, ejecutar un cierto número de cálculos de utilidad esperada, extraordinariamente complejos; no pueden saber quiénes son ni en qué sociedad están, pero eso es casi *todo* lo que no saben. Segundo: en la construcción de la matriz de remuneraciones, los jugadores tienen que adoptar una actitud *optimista* hacia el proceso de regateo, y tienen, por lo tanto, que asignar a cada resultado de estado de naturaleza el valor de ese estado de naturaleza para la persona peor acomodada en él. Tercero: los jugadores tienen que admitir, sin argumentación ni consideración seria, que la adopción oficial de un principio garantiza su afortunada aplicación. Alternativamente, tienen que admitir, también sin argumentación, que el principio de diferencia es uno de los principios que pueden ser, en efecto, afortunadamente aplicados en cualquier estado de naturaleza posible. Cuarto: una vez construida la matriz de remuneraciones según las orientaciones de Rawls, la "maximin" parece inadecuada. Unas simples consideraciones de simetría bastarían para conducir a los jugadores a la coordinación según el principio de diferencia, o a un acuerdo de cooperación según el mismo principio.

Y, por último, lo más importante de todo: el paso decisivo en la argumentación de Rawls, el paso en el que recurre a lo que él llama "maximin", el paso que nosotros aún no hemos sometido a un examen crítico, es la construcción de las casillas no-estado-de-naturaleza en la propia matriz de remuneraciones. Aunque Rawls, en su discusión del tema, pretende estar defendiendo la regla maximin de elección bajo la incertidumbre en las circunstancias especiales de la situación original, en realidad sólo puede ser comprendido como defendiendo una regla particularmente pesimista para derivar la matriz de remuneraciones de la matriz de resultados. Es hora, pues, de que examinemos esa defensa, porque de ella depende cualquier plausibilidad que pueda tener la totalidad de la teoría de Rawls.

LA CONSTRUCCIÓN DE LA MATRIZ DE REMUNERACIONES

Aunque su tratamiento del tema de la "maximin" es largo y complejo (ver Secciones 26-29 de *Teoría de la justicia*), Rawls vacila y titubea de un modo que hace extremadamente difícil admitir sus pretensio-

nes fundamentales. En las páginas iniciales de su discusión, por ejemplo, dice que "es útil como método *heurístico* el pensar en los dos principios como la solución maximin al problema de la justicia social. Existe una relación entre los dos principios y la regla maximin para escoger en condiciones de incertidumbre" (pág. 181, la cursiva es mía). Esto podría inducirnos a creer que Rawls no pretende, realmente, invocar la maximin en el estricto sentido teórico del juego; y, sin embargo, sus argumentos y referencias en las notas a pie de página, en todo el resto de las tres secciones sobre el tema, no contienen tales vacilaciones ni reparos.

Unas páginas después, al explicar sus razones para adoptar la maximin, Rawls niega que esté hablando de la *utilidad* en el sentido que se ha hecho norma en la teoría del juego, en la economía del bienestar y en las disciplinas asociadas. "El punto esencial, sin embargo —escribe—, es el de que, en la justicia como imparcialidad, las partes no conocen su concepción del bien y no pueden estimar su utilidad en el sentido ordinario. En todo caso, nosotros queremos ir más allá de las preferencias *de facto* originadas por unas condiciones dadas. Por lo tanto, las expectativas se basan en un índice de bienes primarios, y las partes hacen su elección en consecuencia". Pero esto, como veremos en seguida, echa a perder, totalmente, dos de sus tres razones para adoptar el método maximin de razonamiento.

Aun a riesgo de parecer que ignoro las explícitas renunciaciones de que Rawls ha llenado su libro, me propongo seguir aquí el mismo curso que he seguido en otras partes de este ensayo. Admitiré que Rawls *está* intentando exponer y defender un argumento formal, y le atribuiré todo el mecanismo formal, las funciones de utilidad y lo demás, que su argumento, sencillamente, presupone.

¿Por qué adoptar una regla tan pesimista para la construcción de la matriz de remuneraciones a partir de la matriz de resultados? En la forma primitiva del modelo, un jugador construya su matriz de remuneraciones por medio de complejos cálculos de utilidad esperada. Una vez introducido el velo de la ignorancia, este curso de acción ya no le está disponible. Como no sabe quién es ni cuál es su particular función de utilidad, no puede estimar la probabilidad de ocupar esta o aquella posición en uno u otro sistema social gobernado por uno de los principios disponibles; tampoco puede determinar a cuánto ascenderían las recompensas de cualquiera de esas posiciones. Hasta ahora, la situación de elección es sencillamente indeterminada, y ni la regla maximin de elección, ni otra alguna, tienen el menor sentido.

Pero Rawls proporciona a los jugadores un conocimiento de los hechos básicos de la sociedad, y esto basta para permitirles construir el tipo de matriz $(m + 1)$ -dimensional de resultados que hemos estado discutiendo. Rawls proporciona también a los jugadores una tosca, pero aprovechable, función de utilidad, bajo la forma de una teoría de planes de vida y de bienes primarios. Pronto veremos que la argumentación de Rawls implica una relación exponencial relativamente "rectangular" entre bienes primarios y utilidad, pero, independientemente de su forma precisa, quedará claro que Rawls *está* comprometido con alguna concepción de la función de utilidad de los jugadores, a pesar de sus renunciaciones. Dice que cada jugador tiene un plan de vida para el que requiere unas cantidades de bienes primarios, por lo que podemos suponer que los jugadores no son indiferentes al volumen de bienes primarios que pueden conseguir. Rawls sugiere que los bienes que no sean primarios tienen una importancia relativamente pequeña para un jugador (porque, en caso contrario, sería irracional que evaluase principios o resultados, puramente en relación con la remuneración en bienes primarios), de donde podemos concluir que la función de utilidad de cada jugador, hablando en líneas generales, es una función del índice de bienes primarios solo. Se nos dice que los bienes primarios son los tipos de cosas de los que uno quiere más bien más que menos, por lo que sabemos que, para cada jugador, la utilidad es una función monótonamente creciente del índice de bienes primarios. Y, naturalmente, como se supone que la teoría de los planes de vida y de los bienes primarios se aplicará indistintamente a todo ser humano, y los jugadores en la situación original no tienen información alguna que los distinga, todos los jugadores tienen idénticamente la misma función de utilidad. Este último punto basta para hacer simétrica la matriz de remuneraciones, y ofrece así la esperanza de una solución.

Pero, una vez dicho todo esto, ¿por qué no permitir a los jugadores que se entreguen a cálculos de utilidad esperada? Naturalmente, los cálculos serían todos idénticos, pero los resultados serían totalmente distintos de los "peores" cálculos dictados por Rawls. Rawls ofrece tres razones para lo que podemos llamar la regla del pesimismo. Como él dice, admitiendo el carácter discutible de la maximin, "las situaciones que hacen plausible esta regla inusitada parecen tener tres rasgos principales" (pág. 182). Estos son:

"(...) la situación tiene que ser tal que resulte imposible el conocimiento de las alternativas posibles, o que al menos sea altamente inseguro." (Pág. 183.)

"(...) la persona que escoge tiene una concepción del bien tal que le importa muy poco o nada lo que pueda ganar todavía por encima del mínimo que seguramente obtiene al seguir la regla maximin. No vale la pena correr algún riesgo para obtener una ventaja ulterior, sobre todo si puede resultar que pierda aquello que realmente tiene valor." (Pág. 183.)

"(...) las alternativas rechazadas tienen resultados que difícilmente pueden aceptarse; la situación implica graves riesgos." (Pág. 183.)

Voy a entregarme a un examen extenso, y, a veces, técnico, de estas tres razones, pero, antes de que nos adentremos en detalles, conviene recordar exactamente lo que se está discutiendo. A pesar de las frecuentes observaciones de Rawls en contrario, la cuestión que tenemos ante nosotros *no* es cuál sería la elección de principios moralmente admirable, u obligatoria, o justa, o recta, para un jugador en la situación original. La cuestión es cuál sería la elección racionalmente egoísta. Se supone que las circunstancias de la elección, incluidas las reglas del juego y la combinación de conocimiento e ignorancia, garantizan que un principio alcanzado a través de la cooperación o de la coordinación por un grupo de jugadores racionalmente egoístas *será, necesariamente*, admirable desde el punto de vista moral, obligatorio, justo y recto. Rawls tiene una tendencia a esfumar este punto, como veremos en el capítulo XVI, pero toda su teoría descansa sobre él, de modo que, indudablemente, debemos atribuírselo.

Consideremos, en primer lugar, la afirmación de Rawls de que un conocimiento de que las alternativas posibles es imposible, o, al menos, altamente inseguro. Su discusión parece sugerir que Rawls ha pensado en la dificultad de decir cómo se produciría una determinada realidad social, y también en la imposibilidad de conocer en qué realidad social están los jugadores y quién es cada jugador en esa sociedad. Pero la atribución que Rawls hace a los jugadores de un conocimiento de los hechos básicos de la sociedad no tiene sentido, en absoluto, a menos que estos *sean* capaces de hacer unos aproximados cálculos de probabilidades de las funciones de una sociedad determinada. ¿Qué otra cosa podría significar la afirmación de que tenían un conocimiento? Lo mismo podría describirse a los cosmonautas perdidos en el espacio diciendo que conocen los "hechos generales de la naturaleza", añadiendo luego que no tenían idea alguna acerca de si resultaría agradable o doloroso el choque contra una estrella.

Evidentemente, lo que Rawls piensa que los jugadores no pueden apreciar es en qué sociedad se encuentran y quiénes son dentro de ella. Ahora bien: la primera de estas cuestiones no es un problema, en lo

que a la construcción de la matriz de remuneraciones se refiere. Como hemos visto, los jugadores tienen que calcular un valor diferente para cada posible estado de sociedad. La posibilidad o la imposibilidad de apreciar en qué estado de sociedad se encuentran volverá a presentarse, ciertamente, una vez construida la matriz, pero, de momento, podemos dejar a un lado ese problema.

En cuanto a la imposibilidad de apreciar quién es uno en una sociedad determinada, Rawls se apoya aquí en las famosas objeciones a lo que, a veces, se llama "el principio de razón insuficiente". La idea es ésta: si quiero determinar la probabilidad de que un dado dé cuatro, y si no tengo idea alguna, en absoluto, de lo que es probable que el dado vaya a dar, puedo sentirme tentado a suponer que, como hay seis resultados posibles (las seis caras del dado), y como yo tengo razón insuficiente para suponer que cualquiera de ellos sea más probable que cualquier otro, la asignación de probabilidad correcta es $1/6$. Pero también puedo suponer que hay cuatro resultados posibles, concretamente un 2, un 4, un 6, o un número impar; con razón insuficiente para asignar a uno de esos resultados una probabilidad más alta que a cualquier otro, concluyo que la probabilidad es $1/4$. También puedo enfocar la situación en el sentido de que en ella saldrá o no saldrá un cuatro, de modo que la probabilidad, siguiendo la misma regla, debe ser $1/2$. Y así sucesivamente. Parece, pues, que la ley de razón insuficiente, como principio para la asignación de probabilidades, conduce a contradicciones.

La reacción usual ante ejemplos de este tipo es la de afirmar: "¡Pero vamos a ver! El modo *natural* de dividir las cosas es el de decir que hay seis resultados, correspondientes a las seis caras. Cualquier otro modo es sencillamente falso e inadecuado." Pero queremos decir eso, solamente porque ya pensamos que sabemos que cada una de las caras tiene la misma probabilidad de salir que cualquier otra. Si *realmente* ignorásemos las probabilidades de los diversos resultados, entonces no tendríamos forma de saber cuál sería la correcta clasificación de resultados. O, dicho de otro modo, nuestra clasificación es, simplemente, una suma de nuestras creencias acerca de la situación, de modo que no podemos dividir los resultados y luego negar, sencillamente, que sabemos cuál de ellos es más probable.

En el juego del regateo, la primera idea de un jugador podría ser la de que es igualmente probable que él será uno cualquiera de los individuos en cualquiera de las sociedades en que él puede terminar. Tal razonamiento le conducirá a algún tipo de cálculo de utilidad esperada.

Pero Rawls rechaza esta jugada. Parece que "en la situación inicial no existen fundamentos objetivos para asumir que cada cual tiene la misma probabilidad que los demás para convertirse en alguien" (pág. 197). Así, unas líneas después, concluye: "Supondré... que las partes descartan todas las posibilidades a las que se llega únicamente a través de este principio. Este supuesto es plausible a la vista de la importancia fundamental que tiene el acuerdo original y del deseo de que nuestra decisión aparezca como responsablemente tomada frente a nuestros descendientes que habrán de verse afectados por ella" (pág. 197). En mi opinión, la referencia a nuestros descendientes es, sencillamente, inoportuna, y la seriedad de la decisión ha sido ya tomada plenamente en cuenta en la construcción de la función de utilidad universal, como veremos. Así se llega a la racionalidad de invocar el principio de razón insuficiente⁴⁹.

Me gustaría indicar que éste es un caso en el que ese principio es especialmente apropiado. Veamos la situación. Un jugador sabe que él es uno de m jugadores, pero no conoce la magnitud del número m (si la conociese, tendría un indicio acerca del carácter de la sociedad de la cual ha llegado). Sabe que esos m jugadores, incluido él mismo, están ya organizados en una sociedad, y que todos los miembros adultos de esa sociedad —y sólo ellos— toman parte en el juego (sería injusto ex-

⁴⁹ John Harsanyi ha señalado que el extremado pesimismo de la regla maximin es matemáticamente equivalente a una determinada asignación de probabilidades a los resultados, concretamente la asignación en la que todo el peso se coloca sobre el resultado de la más baja remuneración. Tal como la teoría de la utilidad se desarrolla habitualmente, se supone que quien ha de adoptar decisiones de un modo racional debe mostrar un comportamiento que obedezca a un conjunto de poderosos, pero plausibles, postulados. Se demuestra entonces que podemos representar ese comportamiento como equivalente a la maximización de la utilidad esperada, siendo las asignaciones de utilidad a los resultados, invariantes respecto a una transformación lineal. Como Harsanyi señala, un hombre que ha de adoptar decisiones, y cuyo comportamiento se ajuste a los postulados, "no puede menos de actuar como si tratase de maximizar su utilidad esperada, calculada sobre la base de cierto conjunto de probabilidades subjetivas" (Harsanyi, *APSR*, Vol. 69, núm. 2 [junio de 1975], 599). Si un participante en el juego de Rawls adopta la maximin, puede construirse como asignando la probabilidad 1 al peor resultado, o puede construirse como asignando el máximo intervalo permitido por su escala de utilidad al vacío existente entre la alternativa peor y la inmediata a la peor; pero no puede construirse como asignando una probabilidad mensurablemente menor que uno a la alternativa peor y también como asignándole un valor conmensurable con el de las otras alternativas. En el tratamiento axiomático de la utilidad, presentar el comportamiento de elección resumido como "la regla de maximin" es, precisamente, hacer así las asignaciones de probabilidad o de valor.

cluir a alguien del juego, o permitir que los visitantes de otra sociedad tomaran parte en él, como diversión). Sabe, además (y esto es fundamental), que la división adecuada de la sociedad que se halla sometida a examen es una división en personas individuales. La estructura moral de Rawls lo garantiza. Así, pues, hay m jugadores; cada uno de ellos es una de las m personas de la sociedad en que viven; la cuestión importante consiste en el modo de distribución de los bienes primarios entre esas m personas; y para cada una de las S sociedades a las que realmente puede pertenecer un jugador, hay una estructura de clase o un modelo de distribución generales, que él conoce (sabe, por ejemplo, que, en una sociedad capitalista industrial avanzada, la estructura de ingresos es piramidal, no inversamente piramidal — y éste es un hecho básico de la sociedad, donde los haya). ¿Qué razón puede tener un jugador para no suponer que él cuenta con una oportunidad igual a $1/m$ de ser cada una de las m personas de su sociedad?

Pasamos luego a la segunda y a la tercera razones del "pesimismo". Tomadas en conjunto, se suman a la pretensión de que la función de utilidad universal, en cuyos términos cada jugador valora los resultados, tiene una forma "rectangular", un tanto similar a la de la curva de la figura 6.

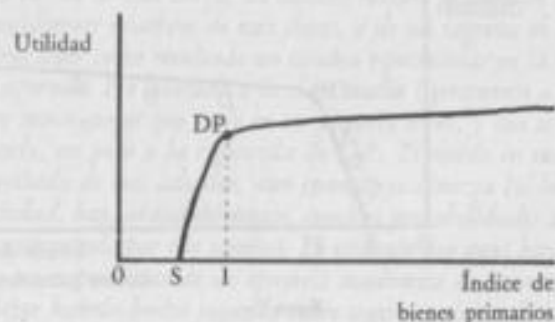


Figura 6

Hay un punto-cerc natural en el eje de utilidad, concretamente el punto correspondiente a una cantidad de bienes primarios, S , claramente insuficiente para sobrevivir. Como los jugadores, según podemos suponer, son indiferentes respecto a los diversos modos en que no logran sobrevivir, este punto es el mínimo absoluto de la curva. A medida que el índice de bienes primarios se eleva, la curva de utilidad se eleva claramente; la vida se hace notablemente mejor. Por último, se

alcanza un punto más allá del cual (teniendo en cuenta todo lo demás, incluida la prioridad de la libertad, etc.) un jugador "se preocupa muy poco, suponiendo que se preocupe algo, de lo que podría ganar". La curva, por lo tanto, se aplanan espectacularmente, aunque sigue ascendiendo, porque los jugadores tienen una utilidad marginal positiva para los bienes primarios. El punto *DP* puede imaginarse como el punto en que el declive de la curva es 1. Podemos fijar, arbitrariamente, la cantidad de bienes primarios correspondiente a ese punto como igual a la unidad.

Supongamos ahora que tuviéramos que ofrecer una jugada a un jugador: la seguridad de una unidad de bienes primarios, que le situará en el punto *DP* de su curva de utilidad, o una probabilidad de 50-50 de una ganancia de q unidades de bienes primarios o una pérdida de r unidades (r tendrá que ser, evidentemente, menor que $1 - S$). Esta es una situación que presenta "graves peligros", como Rawls señala, porque, a juzgar por la forma de la curva, es evidente que una pérdida más bien pequeña, r , de bienes primarios costará al jugador más utilidad que incluso una ganancia muy grande, q .

Pero observemos ahora la figura 7. Si un jugador está en el punto

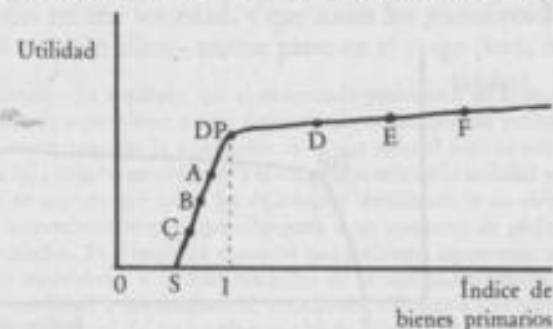


Figura 7

B de su curva de utilidad, puede estar perfectamente dispuesto a entrar en una atractiva jugada de búsqueda entre los puntos *A* y *C*, en lugar de decidirse por la seguridad de *B*. La razón consiste en que el declive de la curva hace que los volúmenes de bienes primarios que debe ganar asciendan de *B* a *A*, aproximadamente en el mismo orden de magnitud en que el volumen que debe perder descende de *B* a *C*. Esto es válido también para una jugada de *D* o de *F* frente a la seguridad de *E*, salvo que, en este caso, hay mucha menos utilidad en juego. Sólo en el punto *DP* y en puntos muy próximos a él, a uno y otro lado, será cier-

to que, como la argumentación de Rawls requiere, la seguridad de *DP* es preferible a toda jugada plausible, no extraordinaria, entre una ganancia, q , y una pérdida, r , de bienes primarios.

Con una curva de utilidad de esta forma, un jugador empieza ahora a construir su matriz de remuneraciones a partir de la matriz de resultados. Si este análisis es correcto, podemos construir un jugador rawlsiano que razone de la manera siguiente:

*Para cada sociedad posible, debo asignar un valor correspondiente a la adopción en esa sociedad de cada uno de los principios disponibles. Empezaré por considerar la Sociedad S_1 . Por mi conocimiento de los hechos básicos de la sociedad, sé que, en una sociedad de ese tipo, el principio 1 tendrá como resultado la formación de un determinado sistema de clases sociales, divididas, aproximadamente, en las siguientes proporciones, y que recibirán, aproximadamente, los siguientes índices de bienes primarios. Como aquí puede utilizarse el principio de razón insuficiente (naturalmente, soy yo quien habla, y no Rawls), puedo efectuar un cálculo de utilidad esperada y obtener un valor que ha de insertarse en la matriz de remuneraciones. Pero esperemos un momento: unas clases de esta sociedad, bajo este principio, caerán a la derecha de *DP* en la curva de utilidad; otras caerán a la izquierda. Dada la forma de esta curva, un cambio incluso pequeño en mi estimación de los volúmenes relativos de esas clases, o de sus ingresos en bienes primarios, podría tener como resultado un cambio espectacular en la suma total de utilidad esperada. He colocado a la clase media ligeramente a la derecha de *DP*. Pero supongamos que estoy en un pequeño error, y que se encuentra, por el contrario, un poco a la izquierda de *DP*. Teniendo en cuenta el carácter improvisado de mis cálculos, aun cuando yo conozca los hechos básicos de la sociedad, hay, indudablemente, muchas probabilidades de que pueda haberme extraviado por este camino. Es evidente que aquí hay graves peligros, y esto no es precisamente un ejercicio académico de econometría — es mi vida. Mejor habría hecho jugando sobre seguro, y eligiendo como el valor que habría de asignarse a este resultado la utilidad esperada por el hombre representativo peor acomodado.*

Mientras el jugador calcule de este modo los valores de los resultados, el principio de diferencia debe proporcionar, por lo menos, una cadena de números (para los diversos estados de sociedad) tan buena como cualquier otro principio. El resultado es una matriz de remuneraciones que, por las razones de simetría discutidas más arriba, inducirá a los jugadores racionales a cooperar o a coordinarse de acuerdo con el principio de diferencia.

Pero, ahora que hemos estudiado la forma de la curva y los detalles del razonamiento, resultará evidente que un jugador actuará según lo que Rawls propone, sólo en cierta combinación limitada de posibles sociedades y principios. Mientras las remuneraciones para un determinado par sociedad/principio parezcan agruparse alrededor de *DP*, de modo que pequeños errores de apreciación produzcan muy grandes cambios en un cálculo de utilidad esperada, puede justificarse una extrema cautela (o pesimismo). Pero si las remuneraciones se sitúan sólo a un lado de *DP*, el razonamiento subyacente en la "maximin" de Rawls se derrumba. Y si la matriz de remuneraciones *no* se construye enteramente a lo largo de las pesimistas líneas de Rawls, *no* puede esperarse que la cooperación o la coordinación conduzcan a la adopción del principio de diferencia. Ya no será necesariamente cierto que, en cada una de las disposiciones asociadas con una realidad social, la casilla correspondiente a la adopción del principio de diferencia sea mayor que cualquier otra.

Como saben todos los lectores de *Teoría de la justicia*, no hay una palabra en el libro que apoye la pretensión de que cualquier sociedad determinada, regida por un principio específico, agrupará sus clases sociales alrededor del punto *DP*. Pero hay algo que está claro: si las remuneraciones *están* agrupadas así para *algunas* combinaciones sociedad/principio, podemos estar seguros de que *no lo estarán* para muchas otras combinaciones sociedad/principio. Por lo tanto, no será racionalmente prudente que un jugador construya la matriz de remuneraciones de la manera supuesta por Rawls; y no hay, en consecuencia, razón alguna para suponer que los participantes en el juego del regateo hayan de coordinarse o cooperar según el principio de diferencia, ni según cualquier otro principio.

Envío técnico

Sólo para dejar bien cerrada la argumentación, volvamos atrás y consideremos el juego del regateo, una vez que la matriz de remuneraciones haya sido construida según las líneas que yo acabo de sugerir, en lugar de serlo a la manera de Rawls. Si aceptamos las implicaciones de la simetría sin disputa, podemos reducir el juego a un simple juego *contra natura*, jugado por un solo jugador (ver pág. 127). La matriz de remuneraciones es bidimensional, y muestra todos los estados sociales posibles a lo largo de una dimensión y todos los principios disponibles (incluida la opción del estado de naturaleza) a lo largo de la otra. Si

hubiéramos construido la matriz de remuneraciones según Rawls prescribe, la simple dominación débil conduciría a la elección del principio de diferencia. Pero si construimos a la matriz de la manera que yo he esbozado, utilizando la regla de pesimismo para combinaciones sociedad/principio cuyas remuneraciones se acumulan en torno al punto *DP* en la curva de utilidad, y haciendo las estimaciones de utilidad esperada de otro modo, entonces, casi seguramente, ningún principio dominará débilmente a ningún otro. En este punto, Rawls puede sentirse tentado a apelar a la maximin en el verdadero sentido, situando el nivel de seguridad de cada hilera y eligiendo luego la estrategia con el nivel de máxima seguridad. Pero esto proporcionará el principio de diferencia, *solo* si la sociedad tecnológicamente más pobre acumula sus remuneraciones en torno a *DP*. En ese caso, la regla de pesimismo dictará que el valor para cada principio será la utilidad del individuo peor acomodado; en tales circunstancias, el principio de diferencia proporcionará el valor más alto; y podemos suponer, para simplificar, que, para cualquier principio, los valores de los estados sociales más ricos serán los más altos. Como el lector puede comprobar trazando, a modo de ejemplo, una matriz de remuneraciones según estas líneas, la maximin conduce, ciertamente, al principio de diferencia. PERO: si algunos de los estados sociales acumulan todas sus remuneraciones a la izquierda de *DP* (un punto —recordémoslo— que se define objetivamente, con independencia de los estados sociales), entonces, muy probablemente, el nivel de seguridad del principio de diferencia *no* será el nivel de máxima seguridad, y algún otro principio será elegido por la maximin. Supongo que, si tuviese que aceptar este análisis, Rawls podría excluir, arbitrariamente, estados sociales cuyas remuneraciones caen a la izquierda de *DP*, aproximadamente igual que Mill, en *Sobre la libertad*, exime al tercer mundo subdesarrollado de las exigencias de la libertad y de la auto-determinación. Pero, como las cuestiones de justicia son particularmente apremiantes a la izquierda de *DP*, ése sería un movimiento desafortunado.

LA FORMA DE LA CURVA DE UTILIDAD Y OTRAS CUESTIONES

¿Por qué hemos de suponer que un participante en la posición original tiene una concepción del bien, en virtud de la cual se preocupa muy poco —en el caso de que se preocupe algo— de lo que puede ganar por encima de un punto *DP*? ¿Es plausible que la función de utilidad típica tenga la forma de las curvas en las figuras 6 y 7? Yo creo

que no. La teoría de los planes de vida y de los bienes primarios requiere que las partes en la posición original tengan objetivos y gustos seculares, no ascéticos. Porque tienen planes de vida racionales, no viven para el momento, saciándose sensualmente sin pensar en el mañana. Ahora, la cuestión consiste en que, en una sociedad como la nuestra, los bienes realmente costosos, los bienes que requieren un alto índice de bienes primarios, son, en su mayor parte, los tipos de bienes que serían deseados por alguien que tuviese un plan racional de vida, más que por un come-bebe-y-sé-feliz cirenaico. Una tripa llena de cerveza y de *prêt-à-porter* requiere muy poco dinero, pero un estilo de vida cultivado, de buen gusto, elegante, racionalmente dirigido a fin de "catalogar actividades, de modo que puedan satisfacerse diversos deseos sin interferencias", cuesta mucho. Si yo supiese que tenía un plan de vida de ese tipo, y si yo conociese los hechos básicos de la sociedad, esperarí que mi curva de utilidad se elevase más bien marcadamente, incluso después de haberse iniciado el aplanamiento. Si se me hubiese asegurado ya una remuneración en alguna parte del sector aplanado de la curva, podría correr, razonablemente, algunos riesgos de caer por debajo de ese punto, a fin de mejorar mis posibilidades de desplazarme hacia la derecha.

Hasta ahora, nos hemos concentrado solamente en la porción de la curva a la derecha del punto *DP*. ¿Qué ocurre con la porción de la izquierda? Concedamos a Rawls, a efectos de la discusión, que la curva se eleva muy bruscamente en esa porción. Aun así, su forma no tiene por qué aterrar a los jugadores hasta el punto de impulsarlos hacia un pesimismo rawlsiano. La razón —aunque resulte bastante extraño— es indicada por el propio Rawls en el curso de una discusión de posibles objeciones a la maximin. Una objeción corriente es el hecho de que el principio de diferencia nos obliga a elegir un curso de acción que garantice una remuneración mínima superior, al margen de que exceda ligeramente los mínimos ofrecidos por cursos de acción alternativos, y al margen de que las otras posibles remuneraciones de las alternativas excedan notablemente las otras remuneraciones del curso de acción elegido. Un poco toscamente: si la acción *A* garantiza \$1, en el peor de los casos, pero ofrece sólo \$1.01 en el mejor, mientras la acción *B* garantiza \$0.99 en el peor y \$1 000.000 en el mejor, entonces la maximin me señala que elija la acción *A*.

A esto, Rawls replica que el principio de diferencia "no está pensado para ser aplicado en tales posibilidades abstractas". Vale la pena citar todo el pasaje:

"Como dije antes, el problema de la justicia social no consiste en distribuir *ad libitum* entre ciertas personas diversas cantidades de algo, sea dinero, propiedades, o cualquier otra cosa. Tampoco existe ninguna substancia componente de las expectativas que pueda transmitirse de un representante a otro en todas las combinaciones posibles. Las posibilidades apuntadas por la objeción no pueden surgir en los casos reales; el conjunto de las posibilidades reales es tan restringido que quedan excluidas. La razón de ello es que los dos principios están vinculados formando una concepción de la justicia que se aplica a la estructura básica de la sociedad como un todo. Precisamente los principios de igual libertad y de justa igualdad de oportunidades impiden que se den estos casos, ya que las expectativas de los más aventajados sólo se elevan en la medida en que ello es necesario para mejorar las de los peor situados. Puesto que las mejores expectativas de los más favorecidos probablemente cubren los costos de su realización o responden a exigencias de la organización, contribuyen con ello al bien común. Aunque no hay ninguna garantía de que las desigualdades no habrán de ser significativas, existe una tendencia persistente a su disminución mediante la creciente disponibilidad de talentos educados y de oportunidades más amplias. Las condiciones establecidas por los otros principios aseguran que las diferencias que probablemente se produzcan serán menores que las que los hombres han tolerado a menudo en el pasado" (185-86).

Puede disculparse al lector por sentir un ligero vértigo ante la suave negativa de Rawls de que en el centro de su teoría se encuentre noción alguna de transmisión de dinero, de propiedades, o de cualquier otra cosa entre hombres representativos. Toda aquella charla acerca de la maximización, todos aquellos gráficos cuasi-económicos, con sus implícitos supuestos de continuidad, divisibilidad, etc., inducirían, ciertamente, al lector desprevenido a tal interpretación. Pero, en lo que a nuestros propósitos se refiere, la clave de este pasaje se encuentra en la pretensión de que las otras dos porciones de los principios, las cláusulas de libertad igual y de igualdad de oportunidades, han excluido un cierto número de patrones de distribución imaginables cuya disponibilidad podría sembrar dudas sobre la racionalidad de la maximin.

El objetivo del pasaje es el de demostrar que los cargos mejor remunerados no estarán fantásticamente mejor remunerados que los peores. Pero el mismo tipo de razonamiento podría conducirnos, con igual facilidad, a la conclusión de que los cargos peor remunerados no serán desastrosamente menos ventajosos sin la maximin que con ella. La libertad igual supone importantes restricciones respecto al grado en que

puede permitirse que esté mal el miembro menos acomodado de una sociedad, así como respecto al grado en que pueden estar bien los que se encuentran en la cumbre. El principio de "cargos abiertos", interpretado por Rawls en el sentido de que significa "justa igualdad de oportunidades", también mejora notablemente la condición de los que se hallan en el extremo más bajo del espectro social.

A la luz de estas consideraciones, hay alguna razón para suponer que el funcionamiento del Principio I y del Principio IIb, por sí solo, excluirá como imposibles aquellos sectores que se encuentran a la izquierda de DP, cuyo temor podría conducir a un jugador a adoptar el modo pesimista de Rawls de valorar los resultados. Al propio tiempo, los argumentos del pasaje antes citado, aunque sugieren, ciertamente, que se excluirían del conjunto de posibilidades espléndidas remuneraciones, atractivas para jugadores naturales, de ningún modo bastan para establecer la pretensión, mucho más fuerte, de que una persona "tiene una concepción del bien tal que le importa muy poco o nada lo que pueda ganar todavía por encima del mínimo que seguramente obtiene al seguir la regla maximin" (183).

Una última cuestión acerca del recurso a la maximin. Un poco más adelante, en la sección titulada "Algunas de las principales razones en favor de los dos principios de la justicia", Rawls recurre a los hechos de la psicología moral, y al problema de estabilidad en la sociedad establecida sobre los principios elegidos en la posición original, para apovar su opción en favor de la maximin. En la elaboración total de su teoría, una de las exigencias que Rawls formula respecto a la elección de principios es la de que no sean principios a los que las partes encuentren difícil adherirse en el mundo real. Más concretamente: las leyes de la psicología humana, disponibles en la posición original, deben estimular a las partes a creer que no habrá excesivas "tensiones de compromiso" para ellas mismas, una vez que reaparezcan en el mundo real y se encuentren viviendo bajo los principios que han elegido. Según esta línea de argumentación, a los que se encuentran en el fondo de la sociedad les costará mucho trabajo aceptar desigualdades que benefician a las clases altas, perjudicándoles a ellos. Rawls señala que, cuando una persona no tiene la sartén por el mango, escaso consuelo encontrará reflexionando que ha jugado y ha perdido. Por ello, surgirá una tensión en el compromiso de las clases inferiores con los principios de la sociedad, tensión que socavaría la estabilidad social¹⁰. Pero, segu-

¹⁰ Aquí, como en otras partes, la argumentación de Rawls presenta una sorprendente semejanza con la argumentación de Platón. Lo que Platón llama "templanza",

ramente, la misma observación puede hacerse respecto a los que surgen del velo para descubrir que se encuentran entre los más capacitados, los más enérgicos y los más productivos miembros de la sociedad. ¿No lamentarán el haber elegido un principio de distribución tan prudente y conservador? ¿Y no creará esto tensiones en su compromiso con ese principio?¹¹.

EL JUEGO DEL REGATEO Y LA JUSTICIA PURAMENTE PROCESAL

Cuando Rawls introduce el velo de la ignorancia, asegura que "la intención de la posición original es establecer un procedimiento equitativo según el cual cualesquiera que sean los principios convenidos, éstos sean justos" (163). Luego explica que "el objetivo es utilizar la noción de la justicia puramente procesal como base de la teoría" (163). El análisis de las distinciones entre justicia procesal imperfecta, perfecta y pura es uno de los muchos y exquisitos fragmentos de razonamiento filosófico de que *Teoría de la justicia* está llena. La cuestión, familiar para todo lector de Rawls, consiste en que, en el caso de la justicia procesal perfecta o imperfecta, tenemos un criterio para determinar el resultado correcto o justo de un procedimiento, independientemente del procedimiento. Podemos juzgar, entonces, si el procedimiento produce siempre, habitualmente, raramente, o nunca, el resultado objetivamente correcto. Por otra parte, en el caso de la justicia procesal pura, la justicia del resultado consiste enteramente en el hecho de que se han empleado procedimientos justos. Muchos resultados alternativos —en realidad, tal vez todos los resultados posibles— pueden, en las debidas circunstancias, ser puramente justos desde el punto de vista procesal, y, en circunstancias indebidas, ser, desde el punto de vista procesal, puramente injustos. El juicio mediante el combate, admitiendo la existencia de un Dios atento, es un ejemplo de justicia procesal perfecta; el juicio mediante jurado, dada la falibilidad del espíritu humano, es un ejemplo de justicia procesal imperfecta; una lotería imparcial es un ejemplo de justicia procesal pura.

Ahora bien: el hecho fundamental en lo que se refiere a la justicia procesal pura consiste en que la imparcialidad está en la realización.

la voluntaria aceptación por parte de cada clase de su función propia en la división social del trabajo, corresponde a la noción de Rawls de la estabilidad y del buen orden.

¹¹ Para una elaborada y detallada discusión del tema, incluido este punto, ver Nozick, *Anarchy, State, and Utopia*, págs. 189 y sigs.

Como Rawls dice: "Un rasgo característico de la justicia puramente procesal es que el procedimiento para determinar el resultado justo tiene que ser efectivamente observado" (109). Si la imparcialidad de los dos principios de la justicia consiste en el hecho de que son el resultado de un juego del regateo puramente justo desde el punto de vista procesal, ese juego del regateo tiene, en efecto, que realizarse, para que los principios sean justos. (O tal vez deberíamos decir para que esos principios sean los principios de la justicia. Hay una diferencia bastante considerable, pero Rawls parece confundir los dos.)

Pero ahora tropezamos con una curiosa paradoja. El resultado de un juego es seguro, y puede ser conocido de antemano, en el caso —y sólo en el caso— de que no haya necesidad de efectuar el juego y de que el resultado no pueda ser descrito como un ejemplo de justicia procesal pura. Si Rawls estuviera proponiendo, por ejemplo, que los puestos desigualmente remunerados de la sociedad fuesen asignados mediante una lotería, resultaría evidente que, para conseguir que la asignación fuese *imparcial*, sería necesario celebrar realmente la lotería. Sería sorprendente que un juez entregase los mejores puestos a sus amigos y que luego justificase su conducta señalando que tal distribución *podía* haber sido el resultado de una vuelta imparcial de la rueda.

Así, pues, Rawls tiene que elegir: o bien puede sostener que el juego del regateo *no* es un ejemplo de justicia procesal pura, o de imparcialidad, y, en ese caso, es libre de tratar de demostrar que se producirá un resultado determinado —la elección del principio de diferencia—; o bien puede sostener que el juego del regateo *es* un ejemplo de justicia procesal pura, y, en ese caso, no puede, en principio, pretender que conoce de antemano cómo se desarrollará el juego.

Creo que está claro que, en realidad, no existe ningún verdadero recurso a la noción de justicia procesal pura en la versión a gran escala que Rawls nos ofrece de su teoría. Esta impresión se ve poderosamente fortalecida por la interpretación kantiana de la posición original, que hemos examinado en la tercera parte. El reino de los fines, tal como se describe en *Fundamentación de la metafísica de las costumbres*, es un agregado de agentes racionales que no se comunican entre sí, de ningún modo esencial. Cada agente quiere, racional y autónomamente, el mismo principio fundamental de moralidad, y no hay, por lo tanto, ningún elemento de justicia procesal pura en la legislación del Imperativo Categórico. Mediante las transformaciones de su teoría, Rawls ha socavado la lógica de su justificación y la ha privado de uno de sus más atractivos rasgos.

Con el extenso razonamiento de esta sección, vuelvo a la afirmación que hice en las secciones iniciales. El corazón de la filosofía de Rawls es la idea del juego del regateo, por medio del cual había de ser superada la esterilidad del razonamiento formal de Kant, y había de establecerse un principio que combinaría las fuerzas y evitaría las debilidades del utilitarismo y del intuicionismo. La idea es original, poderosa y elegante, pero, sencillamente, es insostenible. El esbozo original del juego del regateo era comprensible, pero estaba abierto a aplastantes objeciones. El recurso del velo de la ignorancia permite a Rawls, al menos inicialmente, evitar los escollos del primer modelo, a la vez que parece enlazar su filosofía con la de Kant. Pero el movimiento acaba siendo fatal, porque, al esforzarse por una universalidad absoluta, por una contemplación de los fundamentos de la filosofía social *sub specie aeternitatis*, Rawls hace abstracción de todo lo que es característicamente humano y social. El resultado es un modelo de problema de elección del que no está suficientemente determinado que admita una solución, ni bastante histórico ni humano para sostener una relación útil con las cuestiones reales de la teoría social. En la quinta parte, trataré de decir qué otras vías considero que podrían ser de exploración más provechosa para la teoría social.

XVI

EL STATUS LOGICO DE LA ARGUMENTACION DE RAWLS

SE CUENTA la anécdota del gran medievalista Harry Austryn Wolfson, según la cual, cuando le llegó el momento de retirarse después de casi medio siglo de docencia en Harvard, el Presidente Nathan Pusey le preguntó a quién podría recurrir Harvard para sustituirle. Se dice que Wolfson le respondió: "Bien, en primer lugar, necesitarán ustedes a tres personas". Cuando nos enfrentamos con la pregunta: "¿Cuál es el *status* lógico de la argumentación de Rawls?", lo que se nos ocurre pensar es: "Bien, ante todo, necesitará usted tres respuestas". Rawls intenta unirlas en su doctrina del equilibrio reflexivo, pero la verdad es que, en mi opinión, tiene, realmente, tres respuestas, y no se conjuntan totalmente.

La primera respuesta es la más incitante, la más original, la más audaz, y la más expuesta a crítica¹². Consiste en que los dos principios de la justicia son la solución, en sentido estricto, de un juego del regateo, cuyos términos incorporan una noción mínima de razón práctica, juntamente con las llamadas condiciones de justicia, más la única premisa adicional de que los jugadores están preparados para adoptar un compromiso, de una vez para siempre, con un conjunto de principios para la valoración de prácticas, los principios que han de ser elegidos unánimemente sobre la base del egoísmo racional. Como prueba de esta proposición, "deberíamos aspirar a una especie de geometría moral con todo el rigor que su nombre indica" (146). En virtud de las características del juego del regateo, está garantizado que la solución será justa, de modo que el desarrollo real del juego es un ejemplo de justicia procesal pura. A la pregunta: "¿Por qué debemos gobernar nuestros asuntos de la manera especificada por los principios de la justicia?", podemos contestar, o bien: Porque éstos fueron los principios elegidos en un juego del regateo imparcial por personas (incluido usted mismo) convenientemente situadas, o bien: Porque éstos son los principios que

¹² En vista de la línea que Rawls adopta en uno de sus primeros trabajos, "An Outline of a Decision Procedure in Ethics", que apareció en *Philosophical Review* en 1951, no podemos decir también que fuese la primera respuesta que ideó.

se elegirían en cualquier juego del regateo imparcial, por personas como usted, convenientemente situadas. La fuerza de tal teorema radicaría en sus pocas premisas. Concretamente, la única premisa a la que un egoísta racional (o un devoto de una moralidad competitiva) podría formular una objeción es el acuerdo de adoptar un compromiso obligatorio con un conjunto de principios. Si el teorema pudiera demostrarse, Rawls estaría entonces en condiciones de decir a un crítico: "O reconoce que carece usted totalmente de principios, y que por su negativa a reconocer cualquier principio pierde usted el respeto debido a un ser moral, o, de lo contrario, admite que mis dos principios de la justicia son los principios con los que usted, como agente racional, debe estar dispuesto a comprometerse".

No es un mal desafío para lanzar al rostro de un adversario. Pero —¡ay!— el teorema no puede demostrarse, ni siquiera con los innumerables ajustes, enmiendas, revisiones y apoyaturas que hemos registrado a lo largo de este ensayo. Así, pues, Rawls tiene que pasar, en cambio, a su segunda respuesta, que es una versión un tanto compleja de lo que podría llamarse la "reconstrucción racional" de una conciencia moral ordinaria.

Una reconstrucción racional es una sistematización y un análisis de un cuerpo de convicciones firmemente sostenidas, que revela las conexiones lógicas entre ellas existentes, y las muestra como inferencias, o aplicaciones, o, en cierto sentido, derivadas de algún pequeño conjunto de principios generales. Por ejemplo, el utilitarismo puede interpretarse como una reconstrucción racional de un cuerpo de convicciones morales, destinada a demostrar que todas ellas son incitaciones a elegir la acción que promete producir la mayor felicidad para el mayor número. Así también, la doctrina del medio puede entenderse como una reconstrucción de un cierto número de valoraciones establecidas de personas y de rasgos de carácter, que revela que todas ellas tienen la forma de una alabanza del medio y de una condena de los extremos en unas u otras especies de acción humana.

Una reconstrucción racional no es, en modo alguno, un simple procedimiento, ni está servilmente subordinada a cada detalle de las convicciones morales que constituyen su punto de partida. En el curso de la sistematización de mis convicciones, yo podría descubrir incongruencias o incoherencias insospechadas, y podría, por lo tanto, verme materialmente conducido a cambiar mis creencias. Es, aproximadamente, un proceso de este tipo el que induce a las personas razonables a abandonar prejuicios no examinados.

Un ejemplo bien conocido de la reconstrucción racional en varias fases es la progresión desde el utilitarismo simple, a través del utilitarismo medio, hasta el utilitarismo-norma. En cada caso, las inaceptables implicaciones morales de los principios generales formulados en la fase anterior obligan a una revisión que conduce a una nueva reconstrucción, cuyos principios alcanzan un mejor ajuste con las convicciones de quienes están llevando a cabo la reconstrucción.

Mientras permanecemos en el reino de la reconstrucción racional, el *status* lógico del proyecto es claro, y no excesivamente satisfactorio. La base para afirmar los principios alcanzados por el análisis es, sencillamente, el acuerdo general del auditorio con las convicciones morales originales del autor. En resumen, todo el procedimiento no puede hacer más que aspirar a una versión sofisticada del *consensus gentium*. Huelga decir que una teoría moral de este género no tiene defensa alguna contra el escéptico moral, y ni siquiera tiene un argumento para persuadir a los que, sincera y reflexivamente, rechazan algunas de las convicciones fundamentales en que se basa la reconstrucción.

La tendencia natural, por lo tanto, es la de buscar alguna prueba independiente de los primeros principios una vez que han sido identificados y formulados por medio de la reconstrucción. Probablemente, el mejor ejemplo de este movimiento desde la reconstrucción hasta la prueba se encuentra en *Fundamentación de la metafísica de las costumbres* de Kant. En el capítulo primero, Kant nos dice, explícitamente, que su propósito es el de revelar la presencia del imperativo categórico en los juicios morales ordinarios, que podía esperar que su audiencia compartía con él. En el capítulo segundo, comienza de nuevo con un análisis de la razón volitiva y de la razón práctica, y trata de derivar de ese análisis el mismo principio que ha extraído ya de los supuestos morales en el primer capítulo. Kant comprende, de un modo perfectamente claro, que su empresa carece de valor, a menos que él pueda proporcionar esa derivación.

Desde el comienzo mismo, Rawls se ha visto dividido entre un deseo de deducir los principios de la justicia como solución de su juego del regateo —y, en consecuencia, de establecer que son, en un sentido muy profundo, principios *a priori* de la razón práctica—, y una creencia de que lo más que él puede confiar en conseguir es un análisis de largo alcance de sus convicciones sociales y morales establecidas, que mostrará sus dos principios como una afortunada reconstrucción racional de ellas.

El fuerte argumento *a priori* incorporado en la versión del juego del

regateo de la teoría de Rawls presenta tres importantes dificultades. En primer lugar, como hemos visto, Rawls no puede demostrar el teorema sobre el cual descansa el argumento, y, como él mismo reconoce frecuentemente, los obstáculos con que se tropieza para la demostración de tal teorema son formidables. Pero, en segundo lugar, hay una considerable disputa filosófica precisamente acerca de cuáles son los signos de la "racionalidad práctica". ¿Es la maximización de la utilidad esperada un principio incontrovertible de elección racional bajo riesgo? Difícilmente, si se tiene en cuenta la literatura acerca de la razonabilidad de la aversión al riesgo. ¿Debe un agente racional elegir la maximin, o alguna forma del principio de razón insuficiente, o lo que Leonard Savage llama el principio de "pesar minimax", o uno de los muchos otros principios que han sido propuestos como normas para la elección bajo la incertidumbre?¹³ ¿Es la demanda de ordenamientos completos de alternativas disponibles el más débil criterio imaginable de racionalidad instrumental, como supone la mayor parte de la literatura acerca de la elección racional, o tiene razón A. K. Sen al sugerir que varios ordenamientos parciales y cuasi-ordenamientos reflejan más exactamente nuestras instituciones morales?¹⁴ El conocimiento que Rawls tiene de estas disputas debilita su confianza en la aceptabilidad evidente de toda formalización particular de los principios de la racionalidad formal. En consecuencia, está mucho menos seguro de que su teorema resolviese la cuestión de la justicia social, aun en el caso de que pudiera ser demostrado.

Y, por último, queda la pregunta residual: ¿y qué? Aun cuando pueda lograrse una caracterización satisfactoria de la racionalidad formal, y puedan establecerse los dos principios como la solución, en un sentido convenientemente profundo, de un problema de elección colectiva, ¿por qué ese hecho había de persuadirnos para que adoptásemos los dos principios como los criterios fundamentales para la evaluación o corrección de nuestras instituciones sociales?

Como hemos visto, Rawls revisa sustancialmente su modelo del juego del regateo como respuesta a las dos primeras de estas clases de críticas. Por el momento, se ha impuesto el velo de la ignorancia, juntamente con los supuestos del conocimiento, la sutil teoría del bien, y la noción de los planes de vida; por el momento, los dos principios

¹³ Cf. Luce y Raiffa, *Games and Decisions*, capítulo 13.

¹⁴ Ver Amartya K. Sen, *On Economic Inequality* (Oxford at Clarendon Press, 1973).

han sido convenientemente alterados, tanto por la imposición de normas de prioridad como por la revisión del principio de diferencia, y la norma del pesimismo, extremadamente conservadora, ha sido estipulada como el principio adecuado para la construcción de la matriz de remuneraciones; por el momento, una gran parte de la espectacular plausibilidad inicial de la idea clave de Rawls se ha desvanecido. En cuanto a la tercera objeción, su fuerza real se deriva del hecho de que el modelo de Rawls, aun en su más pura forma, sigue siendo un modelo de heteronomía generalizada, más que de autonomía generalizada. De aquí que, al menos para algunos importantes filósofos morales, no logre captar enteramente la esencia de lo distintivamente moral.

Pero, a pesar de estas graves deficiencias de la versión del juego del regateo de la teoría, Rawls persiste en creer (a mi parecer, con razón) que una afortunada realización de tal proyecto sería una contribución verdaderamente poderosa a la filosofía moral y a la social. Así, pues, resulta claro que no está dispuesto, sencillamente, a renunciar al esfuerzo, y a dedicarse, en cambio, a la reconstrucción racional. El resultado de esta ambivalencia es la característica noción del "equilibrio reflexivo".

La exposición de Rawls del equilibrio reflexivo (Sección 4, "La posición original y su justificación") es elaborada y detallada, y nada se ganaría con resumir aquí sus observaciones. La noción ha sido sometida a un análisis investigador por un buen número de autores¹⁵. La idea central, tal como yo la entiendo, consiste en efectuar una reconstrucción racional que conduzca, no sólo a un conjunto de principios (en este caso, los dos principios de la justicia), sino también a un modelo de elección racional. El modelo tiene que ajustarse luego a la reconstrucción, para que pueda interpretarse, aceptablemente, en el sentido de que proporciona los principios como la solución a un problema de elección planteado en términos de su estructura formal. El propio modelo podrá ser defendido por tener algún mérito independiente como una exposición de los presupuestos de la elección racional. Este mérito, a su vez, presta alguna especie de apoyo al producto de la reconstrucción, además de cualquier apoyo que se derive de la verdad de las convicciones establecidas que fueron el punto de partida de todo el proceso de reconstrucción. El resultado no es tanto una prueba como una disección anatómica del cuerpo de "nuestras" opiniones morales.

¹⁵ Ver especialmente Ronald Dworkin, "The Original Position", reimpresa de la *Chicago Law Review*, en *Reading Rawls*, dirigida por N. Daniels.

El hecho de que los dos principios hayan sido elegidos por las partes en la posición original refuerza nuestra confianza en nuestra comprensión de esos principios, y la profundiza; el hecho de que esos principios hayan sido elegidos en el tipo de posición original eventualmente esbozada por la teoría en su pleno desarrollo fortalece, a su vez, nuestra creencia de que hemos formulado correctamente las condiciones de la posición original. Reconstruir, simplemente, nuestras convicciones en la forma de dos principios para los que no podríamos encontrar ningún apoyo independiente nos dejaría en la incertidumbre de haber articulado nuestras costumbres o nuestros prejuicios. Pero la búsqueda de una prueba *a priori* de esos principios, por medio de una explicación puramente analítica, incontrovertible, de la simple forma de la razón práctica, equivaldría a seguir a Kant en una indagación desesperada. Así, pues, el equilibrio reflexivo debe ser suficiente. Y, evidentemente, Rawls piensa que lo es, porque nada más podría esperarse, racionalmente, de un filósofo moral o de un ciudadano razonable. O eso dice Rawls. Yo sigo creyendo que, en el fondo de su corazón, continúa deseando el teorema, la prueba de unas premisas inalteradas para unas conclusiones innegables.

Pero el equilibrio reflexivo no es más que la segunda respuesta a nuestra pregunta original: ¿cuál es el *status* lógico de la posición de Rawls? La tercera respuesta es más compleja todavía, pues encierra una extensa serie de especulaciones acerca de las instituciones políticas, de las prácticas en la educación infantil y de los ordenamientos sociales, que tiende a apoyar a las partes en su compromiso realista con los principios que se imagina que ellas han elegido en la posición original. Después de ajustar la posición original para acomodar los principios alcanzados mediante la reconstrucción racional, y después de ajustar los principios, a su vez, para hacer de ellos los tipos de principios que las partes en la posición original elegirían, Rawls ajusta ahora los principios y la posición original a fin de obtener unos resultados que, si se llevasen a efecto en el mundo real, estimularían a los hombres a ser los tipos de ciudadanos que cooperarían y sostendrían, de un modo natural y voluntario, las instituciones dictadas por el problema de elección convenientemente dispuesto. El resultado de esta complicación final es el hacer tan complejo el *status* de las pretensiones de Rawls, que yo no estoy seguro, en absoluto, de lo que es premisa y lo que es conclusión, de lo que se defiende con argumentos y lo que se está argumentando para elaborar esa defensa.

Un ejemplo ilustra la dificultad. Se recordará que, ya desde el co-

mienzo, se supone que los jugadores o partes en la posición original no son envidiosos. Como he señalado, el propósito original del supuesto de no envidia es el de hacer posible el recurso a un exceso de desigualdad. Esto constituye una importante restricción en las funciones de utilidad de los jugadores, no exigida por los simples requerimientos formales de la razón práctica. En general, no hay nada irracional en el hecho de preocuparse por la forma en que otros se desenvuelven, y, aunque puede ser reprehensible desear a otros mal en lugar de bien, no parece que eso viole los cánones de la razón práctica. Tenemos, pues, aquí un ejemplo de un ajuste del modelo para hacer que proporcione los principios deseados. Suponiendo que los dos principios estén de acuerdo con nuestras intuiciones acerca de las justas desigualdades de distribución, también podemos interpretar esto como un ejemplo de la propiedad de los principios que conceden un valor a un detalle del modelo.

Pero, en la Sección 80 —una de las últimas—, titulada “El problema de la envidia”, Rawls presenta una línea de argumentación muy diferente. “La razón por la que la envidia plantea un problema” —dice Rawls— es “el hecho de que las desigualdades autorizadas por el principio de la diferencia pueden ser tan grandes que despierten la envidia, en una medida socialmente peligrosa” (587). Debemos determinar —continúa— “si los principios de la justicia, y, en especial, el principio de la diferencia con una justa igualdad de oportunidades, pueden engendrar, en la práctica, una envidia general excesivamente destructiva” (587). Inicialmente, la cuestión de la envidia surgió en relación con la motivación postulada para las partes en la posición original; ahora reaparece, en relación con la cuestión, totalmente distinta, de la armonía y la estabilidad de una sociedad organizada de acuerdo con unas líneas supuestamente justas. Vale la pena de que veamos con detalle la exposición razonada que Rawls dedica a este procedimiento:

“He dividido en dos partes el tema de los principios de la justicia: la primera parte continúa con los supuestos que acabamos de mencionar [que las partes en la posición original no son envidiosas, no tienen actitudes especiales respecto al riesgo, etc.], y se sirve de la mayor parte de lo discutido hasta ahora; la segunda parte pregunta si la sociedad bien ordenada correspondiente a la concepción adoptada generará, realmente, sentimientos de envidia y patrones de actitudes psicológicas que socavarán los ordenamientos que dicha sociedad considera justos. Al principio, razonamos como si no existiesen el problema de la envidia y el de las psicologías especiales; y después, tras haber investigado qué principios se establecerían,

nos detenemos a considerar si las instituciones justas así definidas pueden suscitar y estimular tales propensiones hasta un punto en que el sistema social se hace inviable e incompatible con el bien humano. *Si así fuera, la adopción de la concepción de la justicia debe ser reconsiderada*" (186, la cursiva es mía).

Así, pues, parece que, además de poner de acuerdo la definición de la posición original con los principios alcanzados mediante una reconstrucción racional de nuestras opiniones morales, a través de un proceso de equilibrio reflexivo, también debemos ajustar la una y los otros a nuestras estimaciones empíricas de la estabilidad y de la armonía de una sociedad regida por tales principios.

La lógica de la situación es ahora extremadamente oscura. Consideremos las posibilidades. Supongamos que adoptamos la sólida línea implícita en la concepción original del modelo y afirmamos que todos los principios que surjan del juego del regateo serán, sobre la base de la imparcialidad procesal pura, *justos*. Entonces, podemos argüir, razonablemente, que el conjunto de puntos de solución admisibles en el espacio de regateo está constreñido por consideraciones de posibilidad. Cualquier punto (es decir, cualquier conjunto de principios) que en realidad no pueda ser puesto en práctica será eliminado del juego del regateo, sobre la base de que su adopción significaría, en la práctica, un fracaso a la hora de concertar cualquier solución, en absoluto.

Pero si adoptamos la posición reconstruccionista racional de que los principios de la justicia son la estructura básica implícita de nuestro razonamiento ordinario en cuestiones de moralidad social, entonces el hecho de que una sociedad regida por tales principios pueda ser intrínsecamente inestable podrá ser causa de una desesperación existencial o de una estoica resignación, tal vez, pero difícilmente constituirá una razón para alterar nuestros principios. Después de todo, como el Gran Inquisidor nos recuerda, la verdad es difícil de soportar para los hombres, y una sociedad bien ordenada —una sociedad regida por una concepción públicamente admitida de la justicia— podría no ser un lugar estable ni tranquilo para vivir. Podría haber tentaciones constantes de rechazar la verdad, a cambio de falsas profecías y de consoladoras panaceas pseudo-religiosas.

Pero, finalmente, si demandamos estabilidad como una de las condiciones de la sociedad justa, entonces y sólo entonces tendremos justificación al remontarnos en nuestro razonamiento a unos principios cuya realización promete esa estabilidad, y, más todavía, a un modelo de

elección racional en cuyo interior se contienen supuestos psicológicos que, llevados a la práctica, conducen a una sociedad estable.

En cierto sentido, la argumentación de Rawls es la inversa. *Si* es racional no ser envidioso, y elegir el principio de diferencia en la posición original por medio de una decisión racionalmente egoísta, entonces desearemos un sistema social que reduzca la envidia, porque consideraremos la envidia como irracional. La juzgamos así, porque es mutuamente nociva, al entregarnos a una situación que es Pareto-inferior a otra situación justa y provechosa. Pero, si no es irracional ser envidioso, entonces no rechazaremos necesariamente un sistema social, por otra parte mutuamente beneficioso, que engendra una dolorosa envidia. Podemos considerar la envidia como una emoción legítima, que deseamos fomentar, o que incluso deseamos estimular en los otros.

El *status* lógico de la teoría de Rawls no es claro, en mi opinión, porque, además de su concepción de la elección racional y de sus arraigadas convicciones morales acerca de determinadas cuestiones de justicia social, Rawls tiene también un compromiso extremadamente sólido con una concepción idealista de la sociedad armoniosa y orgánica. Especialmente en las últimas partes de *Teoría de la justicia*, esta concepción es puesta decididamente en juego, y adquiere un peso cada vez mayor en la argumentación. Así, pues, como he observado, hay realmente tres respuestas a la pregunta: ¿cuál es el *status* lógico de la posición de Rawls? La primera respuesta es: Se trata de un esbozo de una demostración de un teorema de la teoría de la elección racional colectiva. La segunda respuesta es: Se trata de un complicado tipo de reconstrucción racional de las convicciones sociales y morales suyas y (así lo espera) de su audiencia, en el que se realiza un ajuste del resultado de la reconstrucción para acomodarlo al modelo de elección racional y para acomodar el modelo de elección racional a él. La tercera respuesta es: Se trata de una visión de un orden social y político armoniosamente integrado, estable, cuya estructura es articulada por los dos principios de la justicia, que, a su vez, son alterados y ajustados a fin de fortalecer la esperanza de que una sociedad vivida bajo su dirección mantendrá, en efecto, su armonía y su estabilidad.

A mi parecer, estas tres respuestas no son compatibles. Las exigencias de cada una socavan los requerimientos de las otras dos. Si la argumentación ha de ser la demostración de un teorema, entonces las premisas no pueden ajustarse indefinidamente del modo más adecuado para proporcionar las consecuencias deseadas. Si la argumentación es una reconstrucción racional de nuestras convicciones morales, enton-

que en lugar del sistema acordado se permitiera elegir política los principios de justicia, y así garantizarlos de estabilidad y de armonía en las relaciones y la cooperación de la cooperación. Pero ahora, en la cooperación de la cooperación, hay una gran dificultad de una teoría de la cooperación, como se ilustra muy claramente en palabras finales del libro de Rawls, respecto a la cooperación. En el capítulo de la cooperación, Rawls muestra que la cooperación es una teoría de la cooperación.

XVII

LA ABSTRACCIÓN DE LA TEORÍA DE RAWLS

¿TIENE razón Rawls? ¿Son *sus* dos principios, realmente, los principios de la justicia? ¿Son éstos los principios de acuerdo con los cuales toda sociedad que muestre las "condiciones de la justicia" *debe* ordenar sus prácticas y distribuir los derechos y las libertades, los bienes y los servicios que sus prácticas colectivas hacen posibles o producen?

Me parece extraordinariamente difícil decidir acerca de esta cuestión, a pesar del cuidado con que Rawls desarrolla temas subsidiarios en su teoría. En parte, el problema surge del hecho de que Rawls dice poco o nada acerca de los hechos concretos de la realidad social, económica y política. Por todas las razones que he expuesto a lo largo de este ensayo, *Teoría de la Justicia* puede situarse históricamente en la tradición de la economía política liberal utópica de finales del siglo XIX y de comienzos del XX. Podríamos caracterizarla brevemente, incluso bruscamente, como una *apología* filosófica de un sello igualitario del capitalismo liberal del estado del bienestar. Y, sin embargo, el recurso del juego del regateo y del velo de la ignorancia, aunque preservando los presupuestos políticos, psicológicos y morales de tal doctrina, suscita la discusión a un nivel de abstracción tan alto, que la especificidad empírica necesaria para prestarle alguna plausibilidad se agota. Lo que queda, en mi opinión, es ideología, es decir, prescripción disfrazada de análisis de valor neutral.

Muchos lectores de *Teoría de la Justicia*, en un esfuerzo por someter las tesis de la obra a la prueba de sus propias intuiciones morales o políticas, han ideado casos de prueba en que la aplicación de los principios pudiera tener importantes implicaciones (es el principio de diferencia —Principio IIa— el que atrae más atención). Inevitablemente, estos casos son más bien de pequeña escala, porque serían necesarios los talentos de un Dickens o de un Tolstoi para crear una sociedad entera como un ejemplo, y los conocimientos de un Weber para idear cómo se desarrollaría esa sociedad.

Como he tratado de demostrar, la formulación original del problema del regateo suscita los casos de pequeña escala o micro-tests; pero,

en su libro, Rawls deja a un lado tales experimentos en el campo del pensamiento, especificando que sus principios deben aplicarse solamente a la estructura básica de una sociedad como conjunto. Así, por ejemplo, debemos ignorar, simplemente, como irrelevantes los tipos de hipotéticos ejemplos de elección, fácilmente representables mediante una matriz de remuneraciones dos-por-dos o tres-por-dos, a cuya prueba podemos someter nuestras intuiciones relativas a la regla maximin para la elección bajo la incertidumbre¹⁶. Algunos críticos se han quejado —yo creo que justificadamente— de que, al excluir los micro-tests en los que nuestras intuiciones son fuertes y la multiplicidad de factores conceptualmente manejables, Rawls ha superprotegido su teoría hasta un punto disparatado¹⁷. De todos modos, estoy de acuerdo con Rawls en que los principios de la justicia social deben concebirse, en primera instancia, aplicados a la sociedad como conjunto, más que a unas prácticas de pequeña escala, de las que puede estar compuesta una sociedad. Por lo tanto, estoy dispuesto a seguir su dirección, y a pensar en los dos principios como aplicados en una dimensión social.

Observemos la sociedad americana contemporánea. Los hechos básicos de la distribución de los ingresos y de la riqueza son bien conocidos y han sido ampliamente discutidos en años recientes. Expuesto en líneas muy generales (porque, para nuestro propósito, es innecesaria la precisión), el quinto inferior de las familias, en términos de ingresos, recibe, aproximadamente, el 5 por 100 de los ingresos distribuidos en un año, mientras el quinto más alto recibe, aproximadamente, el 40 por 100, ocho veces más. Aun teniendo en cuenta hechos atenuantes como el mayor número de "familias" de una sola persona en los quintos inferiores, sigue siendo cierto que los ingresos se hallan muy desigualmente distribuidos. Y la riqueza, desde luego, está mucho más desigualmente distribuida que los ingresos. La mitad más rica del 1 por 100 de todas las familias posee tal vez una cuarta parte de la totalidad de la riqueza, mientras las propiedades de la mitad inferior son despreciables. Los datos sobre los ingresos y la riqueza, tomados en conjunto, incidentalmente, demuestran con toda claridad que es la desigualdad en el ingreso, más que la desigualdad en la riqueza, lo que explica el patrón básico de la distribución en América. Excepto en el caso especial de grandes acumulaciones de capital, que son importantes, sobre

¹⁶ Ver, por ejemplo, Harsanyi, *APSR*, Vol. 69, núm. 2 (junio de 1975), pág. 194. También, pág. 197.

¹⁷ A este respecto, ver Robert Nozick, *Anarchy, State, and Utopia*, págs. 204 y sigs.

todo, como fuentes de poder económico, más que como fuentes de satisfacción privada, lo que más interesa en América es la remuneración que usted percibe por su trabajo, y no el tamaño de su cartera ni la tierra que usted ha heredado¹⁸.

Yo creo que el patrón de distribución de ingresos y riqueza en América es injusto (aunque mi familia y yo seamos en él claramente ganadores, y no perdedores); y el tono de ciertas observaciones accidentales de Rawls sugiere que él está de acuerdo. ¿Qué deberíamos pensar, según él, acerca de esa injusticia? ¿Cómo deberíamos, según él, comparar el actual conjunto de ordenamientos económicos con otros posibles ordenamientos?

Probablemente, tenemos que buscar las expectativas razonables de la familia representativa menos acomodada de nuestra sociedad e indagar si hay algún otro conjunto de ordenamientos que incremente esas expectativas. Ahora bien: yo he dado mucha importancia a la falta de claridad de Rawls respecto al índice de bienes primarios, a la definición del "hombre representativo menos acomodado", a la prioridad lexicográfica de los "bienes" políticos, etc. Pero me propongo ignorar toda ventaja retórica que pudiera conseguirse reactivando esas

¹⁸ Como frecuentemente se oye hablar de millones protegidos por los que se pagan pocos impuestos, y las propuestas de redistribución de ingresos se manipulan siempre de manera que parezca que unas pocas familias ricas son las que nos mantienen sometidos a los doscientos millones restantes, acaso valga la pena señalar algunas de las cifras reales en dólares por las que se hallan separadas entre sí las diferentes fracciones de la población. En 1974, por ejemplo, una mujer y un marido jóvenes, cada uno de ellos recientemente doctorado en Filosofía y con un trabajo de profesor ayudante con \$14 500, podían, explicando un curso de verano, alcanzar justamente el *cinco* por ciento superior de todas las familias americanas (punto de separación: \$31 948). Si un carpintero sindicado, que gana \$7 por hora, y su mujer, que saca como camarera \$2 50 por hora entre salario y propinas, pudieran convencer a su hijo, que aún no ha cumplido los veinte años, de que se gase céspedes y empaquetase comestibles, y si su hija de trece años cuidase niños, regularmente, una tarde por semana, los cuatro se encontrarían en el quinto más alto de las familias americanas (punto de separación: \$20 445). Cuando hablamos de ajustar cuentas a los ricos, no solemos pensar en ninguna de estas familias. A la luz de las cifras reales (que pueden encontrarse, por ejemplo, en el *Statistical Abstract of the United States*), es difícil comprender por qué Rawls atribuye a las partes de la posición original unos planes de vida de tal carácter que prestan poca atención a lo que pueden obtener, en cuanto a bienes primarios, por encima de lo que pueden asegurar mediante el principio de diferencia. Cualquiera que viva hoy en los Estados Unidos tendrá una clara noción de la diferencia en la calidad de vida que se puede comprar, por ejemplo, con \$20 445 por familia, o con \$14 916 (la cima, o punto de separación, del tercer quinto).

críticas aquí, porque yo creo que los problemas de la filosofía social de Rawls son mucho más profundos.

La primera cuestión que debe plantearse es la de saber si hay un exceso indistribuido de desigualdad que está siendo generado en nuestra economía por el conjunto existente de ordenamientos económicos¹⁹. Si suponemos que los plomeros deben ser mejor remunerados que los dependientes de los grandes almacenes, a fin de estimular a los jóvenes (y a las jóvenes, si se impone el Principio IIb) a hacerse plomeros, y también para dedicar a los trabajos de cañería a los que tengan aficiones hidráulicas, ¿deben ser remunerados *más* que los dependientes en la misma medida en que lo son ahora? Hablando en líneas generales, ¿podrían nuestra actual división del trabajo y nuestro patrón de salarios modificarse en el sentido de igualación de las remuneraciones, sin hacerlos inviables? ¿Hay importantes categorías de obreros mejor pagados, cuyos ingresos son mayores de lo que sería necesario para atraerlos, situarlos y mantenerlos en esas ocupaciones mejor remuneradas?

Supongamos que la respuesta a estas preguntas sea —y nosotros sabemos que lo es— sí. Yo creo que lo es, y, en mi opinión, Rawls lo cree también; y como mi querella con él se mantiene en torno a otras cuestiones, me propongo evitar las traidoras arenas movedizas de las estimaciones econométricas.

La segunda cuestión que debemos plantear, siguiendo la dirección de los principios de Rawls, es la de saber si hay otros conjuntos totalmente diferentes de ordenamientos económicos que, aunque sirviendo a los objetivos fundamentales de la producción y la reproducción de bienes y servicios, generarán excesos de desigualdad, cuya posible distribución elevará las expectativas del hombre representativo peor acomodado por encima del nivel que puede alcanzarse en el marco de nuestros actuales ordenamientos mediante la redistribución del exceso de desigualdad existente. Pido perdón al lector por la complejidad de esa exposición, pero la cuestión que la teoría de Rawls exige que nos planteemos es muy complicadamente hipotética. Yo puedo idear un

¹⁹ A fin de evitar el planteamiento de cuestiones metodológicas y teóricas, empleo esta locución neutra, y más bien vaga, de "conjunto de ordenamientos económicos", en lugar de expresiones tan capciosas como "relaciones sociales de producción". Incidentalmente, señalaré que el modo de hablar de Rawls hace abstracción de casi todo lo que es causalmente importante en la vida económica y política de una sociedad.

determinado número de modos distintos de organizar el crecimiento del trigo o el montaje de automóviles —ambas cosas, sin embargo, no son más que ejemplos a pequeña escala o micro-ejemplos, desde el punto de vista de Rawls. Pero, cuando trato de formar una noción utilizable de alternativas a nuestro presente conjunto total de ordenamientos económicos, mi inteligencia no puede hacer nada mejor que ensayar los ordenamientos que realmente han existido en una u otra sociedad —feudalismo, agricultura con trabajo esclavo, caza y acumulación, capitalismo de estado, socialismo colectivista, etc. Después de eliminar aquellos ordenamientos (la esclavitud, por ejemplo) que violan porciones de principios de Rawls distintos del principio de diferencia, tengo que imaginar cómo funcionaría cada uno de los candidatos restantes bajo las condiciones de tecnología, disponibilidad de recursos y nivel de la capacidad actual o potencial de la fuerza de trabajo, que hoy pueden alcanzarse en América. Luego tengo que medir el volumen (si lo hay) del exceso de desigualdad así generado, y valorar el efecto de la redistribución posible más favorable, sobre las expectativas razonables del hombre representativo menos acomodado. Por último, Rawls me dice que clasifique todos los conjuntos de ordenamientos alternativos que se examinan, de acuerdo con la magnitud de las expectativas del menos acomodado. Probablemente, ahora sé qué alternativas son más justas, y cuáles menos justas, que la América de hoy, y el último paso es, sencillamente, señalar en la clasificación el candidato número uno.

La manifiesta vaguedad de estos cálculos y valoraciones tiene una importantísima consecuencia para la teoría de Rawls. Inevitablemente, nos encontramos elaborando el principio de diferencia como un simple principio de distribución. Y, sencillamente, nos detenemos a preguntar cómo se producen, realmente, los bienes que han de ser distribuidos. Robert Nozick da en la diana, muy agudamente, al defender su concepción de los principios de la justicia, frente a la aproximación de Rawls:

"Si las cosas cayesen del cielo como el maná, y nadie tuviese ningún título especial para optar a determinada porción de él [título recibe una significación especial en la teoría de Nozick], y ningún maná cayese a menos que todos estuviesen de acuerdo en una distribución determinada, y la cantidad variase según la distribución, entonces es plausible sostener que las personas situadas de modo que no pudiesen amenazar, ni apoderarse de porciones especialmente grandes, se pusiesen de acuerdo en el principio de diferencia de distribución. Pero, ¿es éste el modelo apropiado para pen-

sar acerca de la forma en que han de distribuirse las cosas que los hombres producen?⁶⁰

Rawls ha respondido a este tipo de crítica. Después de todo —podría decir él—, el principio de diferencia no propone un determinado paquete de bienes y servicios, fijado de una vez para siempre, que ha de ser distribuido. Se aplica a unas prácticas corrientes en las que el esfuerzo, el talento, o cualesquiera otros elementos desiguales están siendo ya legítimamente remunerados de un modo desigual. El principio de diferencia sólo requiere que todo *exceso de desigualdad* sea redistribuido de modo que eleve al máximo las expectativas de los menos acomodados. La teoría tiene plenamente en cuenta —en realidad, debe tener en cuenta— cómo se producen los bienes y los servicios, y cómo respondería la gente que los produce a los cambios que se introdujesen en sus remuneraciones por esa producción.

Pero, aunque Rawls podría responder de ese modo, la crítica de Nozick es muy sólida. El velo de la ignorancia tiene como resultado el de hacer que todas las consideraciones acerca del aspecto de la producción se vuelvan tan enteramente hipotéticas, tan *abstractas* en el mal sentido, que, inevitablemente, el principio de diferencia pasa a ser interpretado como un simple principio de distribución, con los bienes y servicios distribuibles exógenamente dados. No es de extrañar que, cuando otros filósofos tratan de formalizar el principio de diferencia de modo que se le pueda asignar algún tipo de prueba, tiendan a tratarlo de esta manera⁶¹.

Aunque pudiésemos efectuar los cálculos requeridos por los dos principios, no está claro qué significación podríamos atribuir a los resultados, porque los principios no guardan relación, en absoluto, con ninguna noción coherente de causalidad socio-económica. Rawls no

⁶⁰ Nozick, *Anarchy, State, and Utopia*, pág. 198.

⁶¹ Ver, por ejemplo, el tratamiento por Amartya K. Sen, *Collective Choice and Social Welfare* (Holden-Day, 1970), capítulos 9 y 9*. Ver también Steven Strasnick, "Social Choice and the Derivation of Rawls' Difference Principle", *J. Phil.*, 73 (26 feb. 1976), págs. 85-99. Strasnick trata de transformar el juego del regateo en un problema de elección colectiva a lo Arrow, como en cierto sentido hace Sen. Las definiciones de Strasnick son confusas, de modo que, en realidad, su demostración de su teorema no es válida, pero, aunque ordenase las cosas, lo cierto seguiría siendo que su tratamiento transforma el problema en un simple problema de distribución. Ver mi discusión, "Some Remarks Concerning Strasnick's 'Derivation' of Rawls' 'Difference Principle'", programado para aparecer en el *Journal of Philosophy*, en diciembre de 1976.

parece tener concepción alguna de la generación, desarrollo, limitaciones o problemas del poder político. En una palabra, no tiene ninguna teoría del estado.

Cuando se considera que *Teoría de la justicia* es, ante todo, una argumentación en defensa de redistribuciones esenciales de los ingresos y de la riqueza, es asombroso que Rawls preste tan poca atención a los ordenamientos institucionales por medio de los cuales ha de llevarse a cabo la redistribución. No se necesita conocer muchos de los hechos básicos de la sociedad para comprender que se precisaría un poder político muy considerable para imponer los tipos de tasas de salarios, de sistemas contributivos, de pagos de transferencias y de regulación de trabajo exigidos por el principio de diferencia. Los hombres y las mujeres que apliquen el principio, que efectúen los cálculos y dicten las órdenes de redistribución serán las personas más poderosas de la sociedad, ya sean econométricas, representantes elegidos, o reyes de la filosofía. ¿Cómo adquieren este poder? ¿Cómo lo protegen y lo amplían, una vez adquirido? ¿De quiénes son los intereses a que sirven, y de qué modo el servicio de esos intereses los consolidará y los reforzará en relación con otros intereses? ¿Diferirá la organización del poder político, según que las principales acumulaciones de los recursos productivos sean de propiedad privada más que de propiedad colectiva?

Preguntas de este tipo han sido el arsenal del economista político, del sociólogo político y del filósofo político, por lo menos durante los dos siglos transcurridos desde la publicación de *La riqueza de las naciones*, y, de un modo ligeramente distinto, constituyen el esqueleto conceptual de obras clásicas como *La República* de Platón. Liberales del *laissez-faire*, neoclásicos del estado del bienestar y socialistas democráticos discutirán acerca de las respuestas, pero todos estarán de acuerdo en que cualquier teoría de justicia social debe incluir alguna explicación coherente de las fuentes, de la organización, de la distribución y de las funciones del poder político. Que yo sepa, Rawls no tiene tal explicación.

Es fácil encontrar en las páginas de *Teoría de la justicia* dos respuestas a esta imputación. Las prolongadas discusiones de la estabilidad y la congruencia, y las incursiones por la psicología moral abordan, precisamente, la cuestión de la motivación del individuo para cooperar en el sistema de prácticas que se ha establecido bajo la dirección de los dos principios. Y el anuncio explícito, en las primeras páginas, de que la teoría tiene que aplicarse a "sociedades bien ordenadas", soslaya todas las objeciones que mis notas intentan suscitar. Rawls podría decir

que el buen orden es análogo a lo que Platón llamaba la virtud social de la "templanza" en la sociedad justa: es la aceptación voluntaria, por los ciudadanos de la República, de los papeles funcionalmente diferenciados que el principio de la justicia les ha asignado en la buena sociedad. Rawls puede haberse olvidado de evocar para nosotros, como hizo Platón, las formas en las cuales y por las cuales puede degenerar la buena sociedad; pero la teoría platónica no descansa, ni por la evidencia ni por la lógica, en esa brillante interpretación de la decadencia del estado ideal, y la teoría de Rawls tampoco necesita ese apéndice.

Dicho de un modo un poco diferente: Rawls podría argüir que debemos llegar, primero, a una concepción satisfactoria de la forma en que una sociedad *debe* disponer sus prácticas y las correspondientes distribuciones de bienes primarios, antes de que podamos determinar hasta qué punto nuestra sociedad queda lejos de ese ideal y qué pasos debemos dar para rectificar nuestras deficiencias. Sin esa concepción, estaremos siempre confundiendo las disputas sobre principios con desacuerdos en cuanto a los detalles de la reorganización institucional.

Pero respuestas de este género harían caso omiso de otro punto más profundo de la cuestión. La teoría de Rawls de la justicia social es utópica en el sentido en que lo eran las teorías de los primitivos socialistas franceses. Como he dicho ya, su teoría es "abstracta", con lo que quiero decir que hace abstracción de los importantes factores que determinan el carácter y el desarrollo de la realidad social. Esta abstracción se refleja en los modelos de análisis que Rawls emplea. El problema no consiste en que los modelos sean "formales" o en que puedan representarse simbólicamente. Muy al contrario: todo intento de pensar acerca de la realidad social que soslaye esos modelos se quedará al nivel de lo anecdótico. Pero hay modelos buenos y modelos malos, por así decirlo, y los modelos de la teoría del juego, de la teoría del regateo y de la economía del bienestar, en los que Rawls se apoya, son modelos malos para el análisis de la vida social y económica.

Un modelo es como un mapa (o podríamos decir que un mapa es un tipo especial de modelo con el que todos estamos familiarizados). Hace abstracción de la mayor parte de la compleja particularidad de un estado de cosas o de un sistema de objetos y acontecimientos, simplifica y, por consiguiente, agudiza ciertas relaciones, y presenta el resto de un modo que es fácilmente comprendido. El valor de un mapa está en relación con los propósitos para los que se necesita, y lo mismo puede decirse de cualquier modelo.

Supongamos que yo estoy en un bosque (un estado de naturaleza, por así decirlo), y quiero ir a una u otra de las varias ciudades que se encuentran en las cercanías. Naturalmente, necesitaré saber cómo son esas ciudades, a fin de poder decidir hacia cuál de ellas preferiría encaminarme. Pero también necesitaré un mapa que me indique dónde están esas ciudades, en relación con mi situación en el bosque. Ahora bien: un mapa que no me indique más que direcciones y distancias me resultará muy poco útil, porque no me dirá dónde están las cadenas de montañas, ni dónde puedo encontrar pasos transitables a través de ellas. No me advertirá acerca de los ríos, los desiertos y los peligrosos pantanos. Con ese mapa, yo podría suponer, erróneamente, que dos ciudades eran igualmente fáciles de alcanzar, sólo porque se encontraban a iguales distancias desde mi punto de partida. Ciudades muy próximas en el mapa podrían carecer de una ruta que las comunicase directamente, aunque se pudiese llegar a cada una desde el lugar en que yo estaba.

El principio de diferencia de Rawls, en mi opinión, es una regla para componer un "mapa" que tenga todos los defectos que acabo de describir. Es una regla para componer un mapa de indiferencia social, del que se han eliminado todas las características significativas, de orden histórico, político y social. En consecuencia, representa como próximos entre sí unos puntos, cada uno de los cuales puede ser casi inalcanzable desde el otro, y separa otros puntos que, en realidad económica y política, podrían ser virtualmente contiguos⁶².

Veamos los tres siguientes conjuntos posibles de ordenamientos sociales y la forma en que podrían disponerse mediante el principio de diferencia.

El primero es una economía capitalista con un estado poderoso, un gran sector público, muy amplias divergencias de ingresos, y unas instituciones de democracia formal sólidamente establecidas, pero con una considerable desigualdad en la distribución del poder político real entre los más importantes grupos económicos (en resumen, los Estados Unidos de hoy). El segundo es un conjunto de ordenamientos similar al primero, excepto en que el estado, mediante sistemas tributarios y pagos por transferencias, lleva a cabo una redistribución cuyo efecto es, principalmente, el de estrechar las desigualdades de distribución,

⁶² Si le comprendo correctamente, Barber está haciendo casi lo mismo, de un modo un tanto más preciso y técnicamente específico. Ver Barber, "Justifying Justice", págs. 301 y sigs.

un poco el de reducir el producto total y disminuir la tasa de beneficios, pero dejando intacto el capitalismo social, aunque mejorando la suerte de los peor acomodados. El tercero es un estado socialista democrático que tiene el control (y la "propiedad") de las más importantes concentraciones de capital, produciendo, en consecuencia, una distribución de ingresos muy similar a la del segundo (la acrecentada desigualdad de la distribución de la riqueza en el segundo conjunto de ordenamientos podría verse compensada por una mayor amplitud en la estructura de ingresos del tercero).

Estos tres conjuntos de ordenamientos sociales serán, en el modelo compuesto por la teoría de Rawls, tres puntos en el mapa de la indiferencia social. El segundo y el tercero serán preferidos al primero, en virtud de la condición relativa del hombre representativo menos acomodado en cada uno, y, según los he descrito, el segundo y el tercero serán aproximadamente indiferentes (o, para decirlo gramaticalmente, la sociedad será indiferente cuando se le pida que elija entre el segundo y el tercero).

Pero cualquier modelo que represente el segundo y el tercer conjuntos de ordenamientos como "próximos" entre sí no será de utilidad alguna, en absoluto, para valorar las sociedades alternativas, ni para decidir en qué dirección salir de nuestra actual situación. El poder político del capitalismo social cerrará el paso a importantes redistribuciones de ingresos, a no ser que pueda encontrarse algún medio de proteger la situación ventajosa de las sociedades, o bien obteniendo del extranjero un beneficio durante un tiempo de comercio favorable, o bien imponiendo la carga de la redistribución sobre los sectores políticamente inorganizados e ineficaces de los bloques de ingresos medios de la economía (como las pequeñas empresas o los obreros no sindicados, que sufren un descenso real en los ingresos a través de la inflación y del estancamiento en los niveles de los salarios). La alternativa socialista democrática parecerá, naturalmente, mucho más difícil de alcanzar, a un primer examen, pero, mediante el esclarecimiento de los cambios reales en el poder económico que se requieren, podemos enterrar el mito de que la redistribución de ingresos puede alcanzarse sin dolores, a través de la cooperación, armoniosamente, y dentro de la presente estructura de la empresa privada.

Otra forma de llegar a la fuente de la impropiedad de la teoría de Rawls consiste en volver a un punto tocado anteriormente, es decir, a la incapacidad de Rawls de centrarse claramente en la estructura de la producción en la economía, más que en los patrones alternativos de dis-

tribución. Hay una profunda ambivalencia en el pensamiento de Rawls, que recorre toda su caracterización del juego del regateo, su análisis del principio de diferencia, e incluso su psicología moral. Por una parte, como hemos visto ya, Rawls construye toda su teoría sobre la noción de un exceso de desigualdad, lo que requiere alguna concepción de la forma en que se producen los bienes y los servicios, y también —rudimentariamente— de las relaciones sociales en que entran los trabajadores, en el marco de la actividad de la producción. Por otra parte, la idea de un juego del regateo, especialmente de un juego de división equitativa, trata los bienes que han de distribuirse como exógenamente dados. En lo que se refiere a la teoría de los juegos de división equitativa, no existe diferencia alguna entre dividir un pastel que uno u otro de los jugadores ha cocido, y dividir un pastel que ha caído suavemente del cielo. Esto es lo que Nozick quiere decir cuando alude al "maná del cielo". Los modelos económicos utilizados por Rawls muestran el mismo interés por la distribución, con exclusión de la producción. Nada en la noción de la optimalidad de Pareto, ni en el informalismo de un mapa de indiferencia, nos exige que distingamos entre la distribución de bienes y servicios producidos en la diaria reproducción de la vida social, y el reparto de dones gratuitos, descendidos milagrosamente. Podríamos decir que la economía del bienestar es la pura teoría del culto a la consignación.

Ni siquiera toda la panoplia de la economía neoclásica, con sus teorías de la productividad marginal, con sus funciones de producción y con sus teoremas del equilibrio general, nos dice nada acerca de las formas en que la organización de la producción determina la distribución de poder y, en consecuencia, establece patrones sistemáticos de explotación y dominación. Rawls no niega la realidad del poder político, ni pretende que tenga sus raíces fuera de los dispositivos económicos de una sociedad. Pero, al emplear los modelos de análisis de la tradición liberal clásica y de la economía neoclásica, excluye esa realidad de las páginas de su libro. Precisamente porque ha inflado su exposición de una manera desmedida en busca de una integridad sistemática, la ausencia produce el efecto de una negativa, con unas consecuencias que son no simplemente falsas, sino ideológicas.

Y, sin embargo, Rawls es ambivalente, como he dicho. Esa ambivalencia en ninguna parte es más clara que en sus interpretaciones, tanto explícitas como implícitas, de la motivación humana. Como ahora se sabe y se comprende muy bien, la teoría y política clásica —y con ella, los modelos formales de la teoría del juego y de la economía del

bienestar— presupone que los seres humanos buscan la máxima utilidad, persiguen una recompensa y emplean su razón en encontrar la situación más beneficiosa de sus escasos recursos. Nada del agnosticismo psicológico de los tratamientos axiomáticos de la utilidad por von Neumann y Morgenstern y otros puede ocultar la línea directa desde el simple hedonismo egoísta de Bentham hasta las sofisticadas funciones de la utilidad de la teoría del juego y de los mapas de indiferencia de la ritualista economía del bienestar. Al emplear esos modelos de análisis, Rawls participa de sus presupuestos psicológicos subyacentes.

La posición contraria a la concepción utilitarista-liberal de la naturaleza humana es una tradición mucho más antigua, que se remonta hasta Aristóteles y encuentra su más poderosa expresión en los trabajos del joven Marx, según los cuales la actividad creadora, productiva, racional, es el bien para el hombre. El consumo es esencial para la vida; sus satisfacciones forman una componente de la vida buena, cuando se integran adecuadamente en una psiquis sana y bien ordenada. Pero el consumo no es ni puede ser el fin del hombre. Para Marx (aunque no para Aristóteles, salvo en lo que se refiere a la inteligencia), el trabajo de carácter justo es un elemento indispensable de la vida buena. Tratar todo gasto de esfuerzo en tiempo de trabajo como un coste constituye un signo, no de racionalidad prudente, sino de una personalidad desviada y distorsionada. Si todo el trabajo disponible en una sociedad puede ser justamente tratado como un inconveniente por los trabajadores, eso no puede ser interpretado más que como un signo de que las naturales actividades productivas de los seres humanos han sido distorsionadas, en formas tales que destruyen la verdadera felicidad humana.

Ahora bien: lo extraño es que Rawls, más o menos, respalde esta concepción aristotélico-marxista de la naturaleza humana. Incluso invoca el nombre de Aristóteles y sitúa en el centro de su teoría del bien una tesis psicológica a la que da el nombre de "principio aristotélico" (pág. 471). La función peculiar de este principio en la teoría es la de justificar ciertos ordenamientos de paquetes de bienes primarios, desde el punto de vista de la posición original, y no necesitamos entrar en sus detalles. Pero está claro que Rawls no cree estar apoyando ninguna versión actual del hedonismo psicológico del utilitarismo — muy lejos de ello.

Vuelvo, pues, al punto desde el que comencé mi análisis de la filosofía de Rawls. En el corazón de la teoría, en *Teoría de la justicia*, se encuentra un modelo formal de un juego del regateo. El poder de la

teoría consiste en la fuerza creadora y en la imaginación de ese recurso, mediante el cual Rawls esperaba esquivar la estéril disputa entre intuicionismo y utilitarismo. Hablando rigurosamente, desde el interior de la estructura del modo de análisis de Rawls, la maniobra no funcionará porque el modelo, o bien debe atribuir demasiada particularidad a los jugadores, a fin de permitirles regatear con vistas a un resultado determinado y predecible, o, en otro caso, despojarles tan totalmente de sus características distintivas, que no pueda definirse ningún juego del regateo determinado.

Pero, considerado con más amplitud, el fracaso de Rawls procede, natural e inevitablemente, de su acrítica aceptación de los presupuestos socio-políticos y correspondientes modos de análisis de la economía política liberal clásica y neoclásica. Al centrarse exclusivamente en la distribución más que en la producción, Rawls oscurece las raíces reales de esa distribución. Como Marx dice en su *Crítica del Programa de Gotha*:

"Toda distribución de los medios de consumo no es más que una consecuencia de la distribución de las propias condiciones de producción. Esta distribución, sin embargo, es una característica del modo mismo de producción."

¿Tiene razón Rawls? Como sus dos principios de la justicia hacen abstracción de los fundamentos reales de todo orden social y económico, la pregunta no tiene una respuesta útil. ¿Ha buscado Rawls los principios de la justicia, de un modo adecuado? No, porque su teoría, aunque calificada y compleja, es, al fin, una teoría de la pura distribución. La enorme complejidad y la capacidad imaginativa de Rawls nos muestran que el fracaso se debe, no a ningún tipo de insuficiencias de ejecución, sino más bien a la inherente debilidad de toda esa tradición de filosofía política de la que *Teoría de la justicia* es, tal vez, el producto más distinguido.

BIBLIOGRAFÍA

Teoría de la Justicia es el único libro que Rawls ha publicado (Belknap Press of Harvard University Press, 1971. Trad. española: F.C.E., 1979). Todos los ensayos —alrededor de una docena— que ha publicado en revistas académicas tratan de uno u otro aspecto del tema de ese libro. Puede encontrarse una bibliografía de los trabajos y de los artículos de Rawls, hasta comienzos de 1974, en el número especial de *Social Theory and Practice*, vol. 3, núm. 1 (Primavera, 1974), dedicado a *Teoría de la Justicia*. En ese número se incluye el ensayo de Andrew Levine, "Rawls' Kantianism", al que yo me refiero en la tercera parte de este ensayo, así como algunos otros ensayos interesantes.

Reading Rawls: Critical Studies of A Theory of Justice, editado por Norman Daniels (Basic Books, 1974), contiene también una bibliografía de artículos sobre Rawls, algo más completa y más actualizada que la de *Social Theory and Practice*. Entre los muchos ensayos interesantes del libro de Daniels, algunos de los cuales han sido citados aquí en notas a pie de página, se encuentran los de Thomas Nagel, Ronald Dworkin, Gerald Dworkin, T. M. Scanlon, H. L. A. Hart, y Benjamin Barber.

Que yo sepa, sólo hay otro tratamiento de la teoría de Rawls que alcance la extensión de un libro, y es la muy difundida obra de Brian Barry, *The Liberal Theory of Justice: A Critical Examination of the Principal Doctrines in A Theory of Justice by John Rawls* (Oxford, 1973). La aproximación de Barry es algo diferente de la mía, pero yo confío en que los lectores interesados puedan encontrar, tanto en su ensayo como en el mío, mucho material utilizable y muchos motivos de reflexión.

Como contestación a esta afluencia de comentarios acerca de un libro desde cuya aparición escasamente ha transcurrido media década, Rawls ha comenzado a "responder a sus críticos". Además del ensayo, "Fairness to Goodness", en la *Philosophical Review*, citado en una nota a pie de página, ver "Reply to Alexander and Musgrave", *Quarterly Journal of Economics*, 88 (nov. 1974), págs. 633-55.

Uno de los signos de la importancia de Rawls como filósofo político es la insólita amplitud con que estudiosos de campos ajenos a la filosofía han empezado a comentar su teoría. Además de los ensayos de Harsanyi, de Rae y de otros, en la *American Political Science Review* de junio de 1975, he encontrado especialmente sugestivos varios artículos del Profesor Hal R. Varian del M. I. T. Economics Department. Ver, por ejemplo, "Equity, Envy, and Efficiency", en *Journal of Economic Theory*, vol. 9, núm. 1 (sept. 1974), págs. 63-91.

Pero es imposible mantenerse al día, en lo que se refiere a la literatura acerca de Rawls. Cuando estoy escribiendo estas últimas notas, el correo me trae un anuncio de una nueva publicación, *Midwest Studies in Philosophy*, cuyo primer volumen, para febrero de 1976, incluye ensayos de Homer E. Mason, Bernard Baumrin, y Virginia Held, sobre temas directamente relacionados con las cuestiones propias de este libro, y, más extensamente, con el estudio de la filosofía de Rawls. Dejo al autor del próximo libro sobre Rawls la tarea de poner al día la bibliografía.

INDICE ANALITICO

- Acción racional, 20
- Ahorros, tasa adecuada, 89-90, 128
- Apariencia, distinción de Kant entre realidad y, 101
- Aristóteles, 37n., 186
- Arrow, Kenneth, 124n.
- Ascetismo, 123-124
- Asignación aleatoria, principio IIb (puestos abiertos para todos) interpretado como, 48-50
- Autonomy of Reason, The*, 25n.
- Barber, Benjamin, 112n., 125n., 183n.
- Beneficio, maximización, *ver* Maximización del beneficio
- Bien (bondad), 20
 - felicidad y, 21
 - como racionalidad, 71
- Bien ordenada, sociedad, 75-76, 181-182
 - definición, 75
 - envidia y, 169-171
 - status lógico de la argumentación de Rawls y, 169-172
 - publicidad y, 118, 119
- Bienes primarios,
 - índice, 72-73, 82, 83, 123-125
 - teoría, 72-73, 78, 82, 107, 123-124
- Ver también* Ingresos; Libertad; Oportunidades; Poder; Riqueza
- Bienestar, economía del, 40, 45, 65, 185
- Capitalismo, 82, 175
 - principio de diferencia y, 183
- sociedad bien ordenada y, 118-119
- Ciencias naturales, 113-114
- Ciudadanía igual, 77, 82, 83, 84
- Clase media profesional, 126, 127
- Clásica, teoría económica, *ver* Teoría económica clásica
- Clasificación, *ver* Ordenamiento
- Clasificación ordinal, 82n.
- Comparación de prácticas con diferentes números de funciones, 45-46, 63
- Competición,
 - justa, 44, 48
 - Principio IIb (puestos abiertos a todos) y, 48-51
- Compromiso, compulsión del, 23, 24, 26
- Conexión en cadena, 33
- Conocimiento,
 - elemento *a priori* en (Kant), 99-100
 - de la economía, 70, 114-118
 - de probabilidades, regla de pesimismo y, 147-148
 - de psicología, 70, 119-121
 - razón no única fuente de (Kant), 96
 - retenido después de la entrada en el juego del regateo, 65
 - de la sociedad, *ver* Sociedad, conocimiento acerca de la sociología del, 114
 - en la tercera forma del modelo, 69-74
 - Ver también* Velo de la ignorancia
- Constitución política, 77

- Contrato, teorías, *ver* Contrato social, teoría
- Contrato social,
de Locke, 21
de Rousseau, 21-22
teoría (tradiciones contractuales), 19-23, 72, 89
- Contra-transferencia, 120
- Cooperación,
en el juego del regateo, 52, 54, 55, 76
en la primera forma del modelo, 33, 34
en la sociedad, 74
- Coordinación, juego, 54
- Crítica de la razón pura*, 16, 27, 96, 97, 99
- Cuasi-ordenamiento, 33n., 34, 45, 52, 67, 73
- Daniels, Norman, 81n., 112n.
- Democrático, estado socialista, *ver* Estado socialista democrático
- Deseo, 100, 106
- Desigualdad, *ver* Igualdad
- Desigualdad, excedente, 178
en la primera forma del modelo, 35-38, 46-47
- Diferencia, principio, *ver* Principio de diferencia
- Disertación Inaugural*, 96
- Disposiciones económicas de las instituciones,
alternativas a las presentes, 179
importantes, *ver* Instituciones sociales importantes
política diferenciada de las, 76, 83-84
- Distribución, igual o desigual,
el principio de diferencia como principio de, 179-180
ciudadanía igual y desigualdad de, 82
- en la primera forma del modelo, 32, 34-38, 41-52
de ingresos y riquezas en los Estados Unidos, 176-177
exceso de desigualdad y, 35-38, 46-47
libertad y, 40
Principio I (libertad igual) y, 41-42, 44
Principio IIa (las desigualdades actúan en beneficio de todos) y, 43-47
la producción ignorada en favor de, 185, 187
en la economía del bienestar, 43n.
Ver también Ingresos, redistribución; Matriz de remuneraciones; Riqueza, redistribución
- Dominación débil, 135, 136
- Dos principios de la justicia, 24
en la primera forma del modelo, 38, 39, 41-51
interpretación de los dos principios, 41-44
como solución al juego del regateo, 48-52
status lógico de la argumentación de Rawls y, 163-164
concepción micro-económica, 48
respuesta de Rawls a las objeciones contra, 57-59
revisión de la exposición verbal, 58-59
causalidad socio-económica y, 180-181
como solución al juego del regateo, 48-52
improbabilidad de elección de los jugadores, 49-52, 55, 60-61
Ver también Principio I; Principio IIa; Principio IIb
- Dworkin, Ronald, 167n.

- Economía (teoría económica),
clásica, 115-117, 185, 187
conocimiento de, 68, 70, 114, 115, 116-118
macroeconomía, 48, 67
microeconomía, 48
neoclásica, 20, 185, 187
del bienestar, 40, 45, 65, 185
- Eficiencia, ley, 25
- Egoísmo racional,
supuesto, 23-24, 33, 34, 76, 107
en la primera forma del modelo, 33, 76
en el utilitarismo, 19
- Elección (elección racional),
regla maximin de, *ver* Principio maximin de elección bajo la incertidumbre
de tasa de ahorros, 89-90
espacio de soluciones disponibles, definición, 88
Ver también Pareto, optimalidad de; Preferencia unánime
- Envidia, *ver* No-envidia, supuesto
- Equilibrio reflexivo, 163, 167-168
- Estado, 181
liberal burgués, 40
Weber acerca de, 74-
socialista democrático, 184
- Estado-de-naturaleza, opción,
regla maximin de elección y, 135, 136, 144
pesimismo y, 139
como nivel de seguridad, 135-136
realidades sociales y, 142-143
valor, 135-136, 138-139, 144, 145
- Ética, teoría (filosofía moral),
atolladero en, 19-22, 26
aproximación de Kant a, 99-100
- Falsa conciencia colectiva, 115, 116
- Familia monogámica, 77
- Felicidad,
interpretación utilitarista, 19-21
- Filosofía analítica, 16
- Filosofía Crítica de Kant, 97, 99-103
- Filosofía (en general),
enfoque central como valor real de una determinada, 16, 17
concepciones contrarias a, 15-17
moral, *ver* Ética, teoría
- Fines (metas, objetivos, propósitos), 25-26
en la teoría ética de Kant, 99-103
el estado no puede ser definido en términos de, 74
supuestos sustantivos acerca de, ausencia, 65
velo de la ignorancia y, 65-66
- Física, validez *a priori* de, 96, 99-100
- Freud, Sigmund, 119, 121
- Friedman, Milton, 142
- Fundamentación de la metafísica de las costumbres*, 24-26, 99, 102-103, 160, 165
- Geometría, 60
- Gerth, H. H., 75n.
- Gorgias*, 127
- Harsanyi, John C., 67n., 150n., 176n.
- Hart, H. L. A., 81n., 85, 86
- Heteronomía, 106-107, 167
- Hobbes, Thomas, 91
- Hombre representativo,
en la primera forma del modelo, 31, 32, 45
menos aventajado, 44

- principio de diferencia, y, 62-63
 maximización del índice de bienes primarios, 73
 Humanidad, como fin-en-sí-misma (Kant), 102, 103
 Hume, David, 26
- Ignorancia, *ver* Velo de la ignorancia
- Igual ciudadanía, *ver* Ciudadanía igual
- Igual distribución, *ver* Distribución, igual o desigual
- Igualdad, 77, 82, 83, 84
 ante la ley, 42, 82, 85
 de oportunidades, *ver* Principio IIb
 política 82, 85
Ver también Ciudadanía igual de poder, en la primera forma del modelo, 33, 40, 49
- Igualdad política, 82, 85
Ver también Ciudadanía igual
- Imperativo categórico, visión de Kant, 101-103
 interpretación kantiana de Rawls e, 105-108
- Inconsciente, 119-120
- Indicación, instituciones sociales importantes y el problema de, 77
 de bienes primarios, 72-73, 82-83, 123-125
 prioridad de la libertad e, 83-86
- Indiferencia social, mapa principio de diferencia como norma para la construcción, 183-184
- Información, retenida después de entrar en el juego del regateo, 65
- Ver también* Conocimiento; Velo de la ignorancia
- Ingresos, distribución en los Estados Unidos, 176-177
Ver también Distribución, igual o desigual
 problema de indicación e, 124-125
 prioridad de la libertad e, 83, 87
 redistribución, 47-48, 181, 184
- Instituciones sociales importantes (ordenamientos macro-sociales), 67, 74
 como objeto de investigación, 76-77
- Intereses, el velo de la ignorancia en relación con, 65, 73
Ver también Egoísmo racional
- Interpretación kantiana de la posición original, *ver* Kant, Immanuel
- Intuición, 86
 pura, 60
 moral, 19
Ver también Intuicionismo
- Intuicionismo, 19-23, 26, 88, 89
- Juego, teoría, 24n., 34
- Juegos, de dos personas, regla maximin en, 134
 regateo, *ver* Regateo, juego coordinación, 54
 suma-cero, *ver* Suma-cero, juegos
- Justicia, procesal, 37, 159-160
 teoría de Rawls, *ver* Rawls, teoría de la justicia
 dos principios, *ver* Dos principios de la justicia
 "Justicia como imparcialidad"

- (Rawls), 14, 17, 24, 31n., 58
 el juego del regateo en, 131-137
 pesimismo, 136
Ver también Primera forma del modelo
- "Justicia distributiva" (Rawls), 15, 58
Ver también Segunda forma del modelo
- Kant, Immanuel, 16, 20, 24-27, 60, 91, 165
 apariencia-realidad, distinción de, 101
 Filosofía Crítica, 97, 99-103
Crítica de la razón pura, 16, 27, 96, 97, 99
 teoría ética, 99-103
 fundamentación de la metafísica de las costumbres, 24-26, 99, 102, 160, 165
 Rawls y, relación entre, 18, 95-108
 semejanzas, 96-97
 unidad sistemática, 97-98
 su interpretación de la posición original, 86, 88, 95-96, 99, 105-108, 160
- Kierkegaard, Søren, 116
- Laslett, Peter, 15, 58n.
- Levine, Andrew, 105n.
- Ley, autonomía y, 106
 igualdad ante, 42, 82, 85
- Libertad, distribución, en la primera forma del modelo, 41-42
 igual derecho, a la más amplia, *ver* Principio I
 indicación (medida), 84-86
 análisis multidimensional, 85
 prioridad, *ver* Prioridad de libertad
- valor, 86
 de expresión, 85-86
 de pensamiento, 77
- Límite de tiempo, *ver* Juego del regateo y
- Little, I. M. D., 65n.
- Locke, John, 21, 40
- Lógica formal, 16
- Lógica, leyes formales, 25-26
- Lotería, puestos asignados sobre la base de una, 49
- Luce, R. D., 51n., 52, 54, 132n., 143n., 166n.
- Macro-económicos, ordenamientos, *ver* Ordenamientos macro-económicos
- Mannheim, Karl, 127
- Marx, Karl, 46, 114, 186-187
- Matemáticas, 60
 validez *a priori* de,
- Matriz de remuneraciones, construcción de las casillas de no-estado-de-naturaleza en, 145-155
 en "Justicia como imparcialidad", 133, 136
 como simétrica, 138, 145
 bajo el velo de la ignorancia, 137-161
 principio de diferencia, 139-141
- Matriz de resultados, 132-133, 143-144
- Maximin, principio, de elección bajo la incertidumbre, 33n., 50-51, 68, 112, 134
 principio de diferencia y, 136
 matriz de remuneraciones, construcción y, 144-149, 154-158
 pesimismo de, 147-148, 150n.
 psicología y, 158

- argumentación de Rawls en favor de, 78-79
 consideraciones de nivel de seguridad y, 135-136
 opción de estado-de-naturaleza y, 135, 136, 144
 en juegos de suma-cero de dos personas *vs.* juegos de suma-no-cero de *n*-personas, 54-55
 Maximización del beneficio, 128
 Mercado libre, 40
 Mercados competitivos, 77
 Metas, *ver* Fines
 Micro-económicos, ordenamientos, *ver* Ordenamientos micro-económicos
 Mill, John Stuart, 59, 115
 Mills, C. Wright, 73n.
 Moral a Nicomaco, 37n.
 Motivación, 181, 185-186
- Naturaleza,
 conocimiento de la sociedad como diferente del de, 114-117
 estado de, *ver* Estado-de-naturaleza
 humana, 186
 Nivel de seguridad, elección de la estrategia y, 135-136
 No-envidia, supuesto, 169
 en la primera forma del modelo, 34-35, 50n., 51
 excedente de la desigualdad y, 36
 sociedad bien ordenada y, 169-170
 Nozick, Robert, 159n., 176n., 179-180, 185
- Oakeshott, Michael, 128-129
 Objetivos, *ver* Fines
 Obligaciones en el intuicionismo, 19, 20
 Oportunidades,
 igualdad de, *ver* Principio IIb
 prioridad de la libertad y, 83
 Ordenamiento,
 completo, principio de diferencia y, 63, 67
 instituciones sociales importantes y el problema de, 77
 cuasi-ordenamiento de preferencia unánime (optimalidad de Pareto), 33n., 34, 45, 67, 73
Ver también Indicación
 Ordenamientos macro-económicos, 48, 67
 Ordenamientos micro-económicos, 48
 Ordinal, clasificación, *ver* Clasificación ordinal
 Ordinal, preferencia, *ver* Preferencia ordinal
 Pareto, Vilfredo, 33n., 34
 optimalidad de (preferencia o preferibilidad de Pareto), 43
 cuasi-ordenamiento de, 33n., 34, 45, 67, 73
 en la segunda forma del modelo, 57, 58
 fuerte *vs.* débil en la primera forma del modelo, 43-44, 45
Ver también Preferencia unánime
 Paton, H. J., 25n.
 Personalidad: moral, inviolabilidad y dignidad de, 20, 23
 Pesimismo,
 de los participantes en el juego del regateo, 136-139, 141, 143-144
 regla, 147, 155
Philosophy, Politics, and Society, 14n., 15
 Plan racional de vida, *ver* Planes de vida

- Planes de vida, 71-73, 107, 123-129
 concebidos como límites culturales, 126-127
 indicación de bienes primarios y, 123-126
 como tergiversación de la experiencia humana, 127-129
 Oakeshott, acerca de, 128-129
 prioridad de la libertad y, 86
 Rawls, acerca de, 125-126
 visión de la sociedad justa y, 87
 Platón, 114, 127, 158n., 181-182
 Poder,
 igualdad, 33, 40, 49
 político, 181, 185
 Política (instituciones u ordenamientos políticos); la economía diferenciada de, 76-77, 83
 Preferencia,
 ordinal, 33n.
 unánime (unanidad),
 en la primera forma del modelo, 45-46, 50, 51-52, 76
 cuasi-ordenamiento de la relación de, 33n., 34, 46, 52, 67, 73
Ver también Pareto, optimalidad de
 Prichard, H. A., 20n.
Prima facie, obligación, 20
 Primera forma del modelo, 14, 17, 31-55
 los supuestos de fondo de, 39-40
 condiciones, 34, 81-82
 crítica, 39-55
 excedente de la desigualdad en, 35-38, 46-47
 no-envidia, supuesto, en, 34-35, 36, 50n., 51
 prácticas en, 31-33
 Principio I en, 41-43
 aplicación a la evaluación de las prácticas, 44-45
 Principio IIa en, *ver* Principio IIa en la primera forma del modelo
 Principio IIb en, 44, 48, 49, 50, 51
 respuesta de Rawls a las objeciones a, 57-59
 los hombres representativos en, 31, 32, 33, 45
 los dos principios de la justicia en, 38, 39, 41-52
 distribución desigual en, 32, 34-38, 41-52
Ver también "Justicia como imparcialidad"
 Principio de diferencia, 156-157
 estructura básica de la sociedad y, 76
 comparando prácticas con diferentes números de funciones y, 63
 implicaciones contraintuitivas del, 67
 crítica, 66-68
 como principio de distribución, 179-180
 en la primera forma del modelo, 32-33, 48, 57
 utilización, 142, 145
 versión "menos acomodada", 62-63, 73
 maximin y, 136
 matriz de remuneraciones, adopción, 139-141
 alcance, 66, 69
 mapa de indiferencia social según el, 183-184
 supuesto de la "sociedad bien ordenada" y, 141
Ver también Principio IIa

- Principio de puestos abiertos, *ver* Principio IIIb
- Principio maximin de elección bajo la incertidumbre, *ver* Maximin, principio, de elección bajo la incertidumbre
- Principio I (la más amplia libertad compatible con una libertad igual para todos), 82
como distribución igual, no principio de libertad igual, 42, 45
en la primera forma del modelo, 41-43
aplicación a la evaluación de prácticas, 44
problemas de indicación (medida) planteados por, 84-86
maximin y, 156, 157
significado y aplicación de, 84, 86
prioridad, 83, 84
Ver también Prioridad de la libertad
improbabilidad de elección de las partes, 86
valor de la libertad y, 86
Ver también Dos principios de la justicia
- Principio IIa (las desigualdades actúan en beneficio de todos), comparación de las prácticas con diferente número de funciones y, 45, 46
en la primera forma del modelo, 43-44
alteración en las prácticas existentes, 46-48
aplicación del Principio, 44-48
preferencia unánime, 46
en la segunda forma del modelo, 58, 62
Ver también Principio de diferencia: Dos principios de la justicia
- Principio IIIb (puestos abiertos para todos)
versión competitiva *vs.* versión de lotería de, 48-49
en la primera fase del modelo, 44, 48-51, 52
regla maximin y, 156, 157
interpretaciones cambiantes posibles, 44, 48-49
Ver también Dos principios de la justicia
- Principios morales (principios de razón práctica), aproximación de Kant a, 99-101
Ver también Imperativo categórico
- Prioridad de la libertad, 17, 41, 58, 77, 81-87
ingreso y riqueza y, 82, 83, 87
problemas de indicación y, 82-86
interpretación kantiana de la posición original y, 86, 88, 108
nivel mínimo de bienestar requerido, 84, 89
valor de la libertad y, 86
- Producción, incapacidad de Rawls para centrarse en, 184-185, 187
- Propiedad privada de los medios de producción, 77, 118
- Propósitos *ver* Fines
- Prudencia, Ley, 25-26
- Psicoanálisis, 119-121
- Psicología,
conocimiento, 70, 119-121
regla del máximo y, 158
moral, 35
- Psicología moral, 35
- Publicidad,
en el juego del regateo (primera forma del modelo), 40

- sociedad bien ordenada y, 118-119
- Punto de vista moral,
elección de la tasa de ahorro y, 90
velo de la ignorancia y, 61-62
- Racionalistas, Oakeshott acerca del espíritu de, 128-129
- Rae, Douglas, 67n.
- Raiffa, H., 51n., 52, 54, 132n., 143n., 166n.
- Rawls, teoría de la justicia (modelo de Rawls)
abstracción de, 175-187
reconstrucción histórica del desarrollo de, 14-15
clave (idea central) de, 17, 23-27
status lógico, 163-172
organización de este ensayo acerca de, 17-18
problema abordado por, 19-22
fases en el desarrollo de, 14-15
Ver también Primera forma del modelo; Segunda forma del modelo; Tercera forma del modelo
unidad sistemática de, 97-98
como utopía, 175, 181
totalidad de, 79
Ver también Teoría de la justicia
- Razón insuficiente, principio, 149, 150
- Razón práctica, 96, 97, 165-166
leyes formales, 25
intuicionismo y, 19-20
principio, aproximación de Kant a, 99-101
Ver también Imperativo categórico
principios morales sustantivos derivados de los principios formales de (Kant), 102-103
- Razón (razonamiento), interpretación de Kant, 95-96
práctica, *ver* Razón práctica
prudencial, 61, 71, 128
teoría ética de Kant y, 99-102
teórica, interpretación de Kant, 98-100
Ver también Lógica, leyes formales
- Razón teórica (o conocimiento), interpretación de Kant, 98-100
- Razonamiento prudencial (o juicios prudenciales), 61, 71, 128
teoría ética de Kant y, 99-101
- Realidad social, estructura del juego del regateo y, 142-143
- Recompensas,
en la primera forma del modelo, 31, 32, 34, 35, 38
Ver también Distribución; Matriz de remuneraciones
- Reconstrucción racional de opiniones morales, 164-166, 167-168
- Recto, lo (rectitud), 21
en el intuicionismo, 20
- Redistribución de los ingresos y la riqueza, 47-48, 181, 184-185
- "Refutación del teorema de la justicia de Rawls", 31n.
- Regateo, juego, 23, 24, 25, 26, 69, 70
objetivo, 24
características de los jugadores de, 33-34, 39-40, 75-76
condiciones, 33-34, 82
compulsión del compromiso en, 23, 24
como premisa procesal cuasi-formal, 23, 26
cooperación en, 52, 54, 55, 76
análisis formal, 131, 161
estructura del juego, 131-145
utilización de los principios adop-

- tados en, 141, 145, 170
información retenida después de entrar, 65
Ver también Conocimiento;
Velo de la ignorancia
interpretación kantiana, *ver* Kant,
Immanuel, su interpretación
de la posición original
como juego cooperativo de suma
no-cero de n -personas, 52,
54-55
matriz de remuneraciones en, *ver*
Matriz de remuneraciones
prácticas,
concepto, 24
definición, 31
en la primera forma del mode-
lo, 31-32
patrones de recompensa de,
31, 32
publicidad, 40, 118, 119
justicia procesal pura y, 159-160
límite de tiempo y, 53
reglas, 52-53
compulsiones de reglas proce-
sales, 96-97
solución a, 23, 31, 61, 111-112
espacio de posibles puntos de
solución, 88
dos principios de la justicia,
48-51
velo de la ignorancia, 73
estructura de, 131-145
simetría y, 40, 53-54
velo de la ignorancia en, *ver* Velo
de la ignorancia
Ver también Rawls, teoría de la
justicia; y temas específicos
Religión, 72, 116
Remuneraciones, *ver* Recompensas;
Distribución desigual
República, *La*, 114, 158n.
Riesgo, aversión, 55
Riqueza,
distribución en los Estados Uni-
dos, 176-177
Ver también Distribución
prioridad de la libertad y, 82,
83, 87
redistribución, 181
Ross, W. D., 20
Rousseau, Jean-Jacques, 21-22, 87
Runciman, W. G., 15
Savage, Leonard, 166
Schelling, Thomas, 53n., 54
Segunda forma del modelo, 14-15,
17, 57-68
crítica, 65-68
principio de diferencia en, 62-63
velo de la ignorancia en, 60-61
Segundo Tratado, 21
Sen, Amartya K., 166, 180n.
Sentidos, los (sensibilidad), interpre-
tación de Kant de, 96, 105
Simetría, juego del regateo y, *ver* Re-
gateo, juego, simetría y
Soberanía en el contrato social, teo-
rias, 21
Ver también Contrato social
Social Choice and Individual Values,
14
Socialismo, 184
Sociedad (organización social),
según la definición de Rawls, 73-
75
principio de diferencia y estructu-
ra básica, 76-77
justa, visión de Rawls, 86-87
conocimiento de las partes acerca
de (en la tercera forma del
modelo), 69-71, 73-74, 106-
107, 111-121
naturaleza ahistórica (intempo-
ral) de, 114
diferente del conocimiento de
la naturaleza, 114-117

- sociedad bien ordenada, con-
cepto, 118-119
reglas, 74
leyes, 115-116
tasa de ahorro, 89-91
estable, *ver* Bien ordenada, socie-
dad
bien ordenada, *ver* Bien ordenada,
sociedad
Sociología,
funcionalista, 49
del conocimiento, 114
Solución al juego del regateo, *ver*
Regateo, juego, solución a
Strasnick, Steven, 180n.
Suma-cero, juegos,
el juego del regateo como coope-
rativa de suma no-cero de n -
personas, 52
regla maximin de elección y, 54-
55
dos personas, 54-55
Teitelman, Michael, 123n.
Teorema, 16
propósito de Rawls de demostrar
un, 13-14, 23
Teoría económica clásica, 115-117,
185, 187
Teoría ética, *ver* Ética, teoría
Teoría de la justicia, 35, 59, 91-92
complejidad, 13, 14
insuficiencias y ambigüedades,
13-14
propósitos teorematizados, 13-15
estilo de crítica e interpretación
de Wolff, 15
Ver también Rawls, teoría de la
justicia; Tercera forma del
modelo
Teoría de la justicia de Rawls, *ver*
Rawls, teoría de la justicia
Tercera forma del modelo, 15, 17,
69-92
problemas de indicación y orde-
namiento en, 77
conocimiento de las partes en,
69-73
planes de vida en, *ver* Planes de
vida
instituciones sociales importantes
en, *ver* Instituciones sociales
importantes
regla maximin en, *ver* Maximin,
principio, de elección bajo la
incertidumbre
bienes primarios en, *ver* Bienes
primarios
sociedad definida en, 74-75
sociedad bien ordenada como su-
puesto de, 75-76
Transferencia, 119-120
Unidad sistemática, 97-98
Universalidad, autonomía y, 106
Utilidad
media, principio, 55
Principio IIa y, 51
cardinal, 33n., 82n.
esperada, cálculos, y matriz de re-
muneraciones, 146-150
forma de la curva, 151-153,
155-157
Utilidad, funciones (principios de
utilidad)
media, 55
Principio IIa y, 51
cardinal, 33n., 82n.
regla maximin de elección y, 54-
55
matriz de remuneraciones y, 146-
147
universal, 150-152
Utilitarismo, 26, 65, 88-89, 164-
165
intuicionismo y, 19-21, 22
versiones "de noble mentira" de,
76

visión de Rawls acerca de, 20-21
virtudes y defectos, 19
supuesto de la sociedad bien ordenada y, 76

Varian, Hal R., 190

Velo de la ignorancia, 48, 54, 57, 78, 91, 99, 137-144

sus atractivos desde el punto de vista de Rawls, 60-62
crítica, 65-66

condiciones del conocimiento en la posición original y, 112-113, 118, 119

punto de vista moral y, 61-62

matriz de remuneraciones bajo, 137-161

principio de diferencia, 139-141
problemas planteados por, 65-66, 69

Rawls acerca de, 60

Ver también Kant, su interpretación de la posición original

Voluntad,

general (Rousseau), 21-22

privada (Rousseau), 22

autónoma vs. heterónoma, 106

Weber, Max, 74

Wolff, Cynthia Griffin, 119n.

Wolfson, Harry Austryn, 163

INDICE GENERAL

Prefacio	9
----------------	---

Primera Parte

INTRODUCCION

I. <i>Observaciones preliminares</i>	13
II. <i>El problema</i>	19
III. <i>La clave</i>	23

Segunda Parte

EL DESARROLLO DE LA TEORIA

IV. <i>La primera forma del modelo</i>	31
V. <i>Una crítica de la primera forma del modelo</i>	39
Los supuestos de fondo, 39; La interpretación de los dos principios, 41; La aplicación de los principios, 44; ¿Son los dos principios la solución del juego del regateo?, 48.	
VI. <i>La segunda forma del modelo</i>	57
El velo de la ignorancia, 60; El principio de diferencia, 62.	
VII. <i>Crítica de la segunda forma del modelo</i>	65
VIII. <i>La tercera forma del modelo</i>	69
IX. <i>La prioridad de la libertad y otras complicaciones</i> ...	81
La prioridad de la libertad, 81; Otras complicaciones, 88.	

Tercera Parte
RAWLS Y KANT

X. <i>Kant y Rawls</i>	95
XI. <i>El antecedente kantiano</i>	99
XII. <i>La interpretación kantiana de la posición original</i> ...	105

Cuarta Parte
CRITICA DE LA TEORIA

XIII. <i>Los hechos generales acerca de la sociedad humana</i> ..	111
XIV. <i>Bienes primarios y planes de vida</i>	123
XV. <i>Un análisis formal del juego del regateo</i>	131
Estructura del juego, 131; La construcción de la matriz de remuneraciones, 145; La forma de la curva de utilidad y otras cuestiones, 155; El juego del regateo y la justicia puramente procesal, 159.	
XVI. <i>El status lógico de la argumentación de Rawls</i>	163

Quinta Parte
¿TIENE RAZON RAWLS?

XVII. <i>La abstracción de la teoría de Rawls</i>	175
<i>Bibliografía</i>	189
<i>Índice analítico</i>	191

Este libro se terminó de imprimir el día 30 de junio de 1981 en los talleres EDIMEX, S. A., Calle 3 núm. 9, Alce Blanco, Naulcalpan, Estado de México. Se tiraron 3 000 ejemplares.